



How well do different tracers constrain the firn diffusivity profile?

C. M. Trudinger¹, I. G. Enting², P. J. Rayner³, D. M. Etheridge¹, C. Buizert^{4,5}, M. Rubino^{1,4}, P. B. Krummel¹, and T. Blunier⁴

¹Centre for Australian Weather and Climate Research, CSIRO Marine and Atmospheric Research, Aspendale, Victoria, Australia

²MASCOS, University of Melbourne, 3010, Victoria, Australia

³School of Earth Sciences, University of Melbourne, 3010, Victoria, Australia

⁴Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Juliane Maries vej 30, 2100 Copenhagen, Denmark

⁵College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis OR 97331, USA

Correspondence to: C. M. Trudinger (cathy.trudinger@csiro.au)

Received: 24 June 2012 – Published in Atmos. Chem. Phys. Discuss.: 19 July 2012

Revised: 28 December 2012 – Accepted: 22 January 2013 – Published: 6 February 2013

Abstract. Firn air transport models are used to interpret measurements of the composition of air in firn and bubbles trapped in ice in order to reconstruct past atmospheric composition. The diffusivity profile in the firn is usually calibrated by comparing modelled and measured concentrations for tracers with known atmospheric history. However, in most cases this is an under-determined inverse problem, often with multiple solutions giving an adequate fit to the data (this is known as equifinality). Here we describe a method to estimate the firn diffusivity profile that allows multiple solutions to be identified, in order to quantify the uncertainty in diffusivity due to equifinality. We then look at how well different combinations of tracers constrain the firn diffusivity profile. Tracers with rapid atmospheric variations like CH_3CCl_3 , HFCs and $^{14}\text{CO}_2$ are most useful for constraining molecular diffusivity, while $\delta^{15}\text{N}_2$ is useful for constraining parameters related to convective mixing near the surface. When errors in the observations are small and Gaussian, three carefully selected tracers are able to constrain the molecular diffusivity profile well with minimal equifinality. However, with realistic data errors or additional processes to constrain, there is benefit to including as many tracers as possible to reduce the uncertainties. We calculate CO_2 age distributions and their spectral widths with uncertainties for five firn sites (NEEM, DE08-2, DSSW20K, South Pole 1995 and South Pole 2001) with quite different characteristics and tracers available for calibration. We recommend moving away from the use of a firn model with one calibrated parameter

set to infer atmospheric histories, and instead suggest using multiple parameter sets, preferably with multiple representations of uncertain processes, to assist in quantification of the uncertainties.

1 Introduction

Firn is the porous layer of compacted snow overlying an ice sheet. Air is contained in the open pores, and its composition is influenced by changes in the composition of the overlying atmosphere and by processes that occur in the firn. Air can be extracted from the firn and its composition measured (Schwander et al., 1993), providing an archive of old air from which to deduce the atmospheric histories and budgets of trace gases. Firn is also the medium that is traversed by air before being trapped into bubbles in ice. It is for both of these reasons that we are interested in the firn processes, which include advection downwards as new snow falls at the surface, convective mixing due to wind pumping near the surface (Colbeck, 1989; Severinghaus et al., 2001; Kawamura et al., 2006), molecular diffusion through the firn column (Schwander et al., 1988), enrichment of heavier molecules with depth due to gravitational settling (Craig et al., 1988; Schwander, 1989), thermal fractionation due to temperature gradients (Severinghaus et al., 2001), upward flow of air due to compression (Rommelaere et al., 1997) and gradual trapping of air into bubbles (Schwander et al., 1988). Additional

processes that are not as well understood and not widely accepted to be important include bubble close-off fractionation for smaller molecules (Huber et al., 2006), in situ production (Anklin et al., 1995; van der Kemp et al., 2000; Smith et al., 2000) and possible dispersive mixing at the bottom of the firn (Severinghaus et al., 2010; Buizert et al., 2012).

Numerical models of the firn processes are important for interpreting firn and ice core measurements. The rate of diffusion of trace gases through the firn is a poorly known key model parameter, and consequently firn air models require calibration of the depth profile of effective diffusivity for each site. For brevity we will refer to effective diffusivity simply as “diffusivity”. It is possible to measure the diffusivity in small firn samples (Schwander et al., 1988), however Fabre et al. (2000) concluded that variations in porosity would need to be known continuously and precisely over a larger scale than a core sample, particularly near pore close-off, to reflect the macro scale variations in diffusivity and therefore be useful in firn models. Although there have been significant advances recently in characterizing the 3-dimensional structure of firn (Freitag et al., 2002, 2004), it remains to be seen whether lateral heterogeneities will limit the applicability of measurements along a core for representing the diffusivity in situ in the firn. Therefore, the approach of tuning the one-dimensional diffusivity by trying to match modelled and observed mixing ratio profiles in the firn for one or more tracers with known atmospheric history (reference tracers) currently remains the most common, as well as the most accurate, way to estimate firn diffusivity.

Most studies have calibrated firn diffusivity using one or a few tracers, either manually (e.g. Trudinger et al., 1997) or with an automated calibration method (e.g. Rommelaere et al., 1997; Trudinger et al., 2002). Two recent studies used up to 10 tracers at once (Buizert et al., 2012; Witrant et al., 2012). In most cases a single diffusivity profile is estimated without an uncertainty range, although Rommelaere et al. (1997) and Fabre et al. (2000) estimated the uncertainty by perturbing the concentration observations, and Buizert et al. (2012) compared diffusivity determined by six different firn models using the same dataset and physical firn characteristics.

There are three main types of uncertainty in model predictions (e.g. Højberg and Refsgaard, 2005). The first is related to data uncertainties, including calibration data and input data. Data uncertainties can be difficult to quantify, and when errors are systematic rather than random (e.g. calibration biases between firn concentrations and the atmospheric history of reference tracers) this can be difficult to diagnose, providing a challenge for optimisation methods based on assumptions of random noise. The second type of uncertainty is related to model parameters, including the fact that when a problem is under-determined there will usually be multiple solutions that give an adequate match to observations – this is sometimes called equifinality (Beven, 2006) or the null space in linear algebra. The problem of tuning firn diffusiv-

ity using concentration profiles is often thought of as under-determined (Rommelaere et al., 1997; Buizert et al., 2012). While we have generally assumed that as more tracers with different types of atmospheric history are used to tune the diffusivity profile it will be constrained more tightly, to our knowledge this assumption has not been tested or quantified. The third type of uncertainty, which is by far the most difficult to quantify but could be the most important uncertainty, is to do with model error, including the choice of processes included in the model and their mathematical formulations. This type of uncertainty is about how well the model processes represent the real world, and in firn models would include the simplifying assumption of one-dimensional diffusivity, as well as uncertain or unknown processes. All three types of uncertainty are potentially important for firn modelling, but they are often not all taken into account when firn models are used to reconstruct atmospheric records of trace gases. The importance of the uncertainties will depend on how the model is calibrated and what it is being used for. Our focus here will be mainly on the first two types of uncertainty, but we will touch on the third type. In particular, we wish to challenge the idea of using a firn model with only a single diffusivity profile, instead we suggest the use of multiple parameter sets to represent these uncertainties, as has become popular in other fields such as hydrology (e.g. Vrugt et al., 2003).

Here we describe an updated version of the CSIRO firn model (Trudinger et al., 1997) and use it to explore how well different combinations of tracers constrain the diffusivity profile and other parameters related to mixing. We use a global search method to locate multiple solutions that match the observations. We initially use synthetic observations (with errors) based on the NEEM site in Greenland, to test our methodology for cases where we know the answer (i.e. the diffusivity profile that was used to create the synthetic observations), and also to look at what types of atmospheric history provide the best constraints. We then use real observations from a number of sites (NEEM as well as the Antarctic sites DE08-2, DSSW20K and South Pole). We provide a way to represent the uncertainty due to equifinality, to determine the consequences of equifinality for inferring atmospheric histories of other trace gases in the firn.

2 Methods

In this section, we will describe the updated CSIRO firn model, our method for estimating diffusivity, the real and synthetic observations and other model input variables used.

2.1 CSIRO firn model

The CSIRO firn model is based on the model described by Trudinger et al. (1997) and Trudinger et al. (2002), but has been substantially improved and rewritten in Fortran90.

Major changes from the previous version are (i) the inclusion of the upward flux of air due to compression of pore space that we had previously neglected, (ii) improved mass conservation of air in bubble trapping, (iii) a genetic algorithm (GA) has been added to optimise diffusivity and other parameters, (iv) an implicit time stepping scheme as in Rommelaere et al. (1997) has replaced the Euler predictor-corrector scheme (making a larger time step possible), (v) flux smoothing that was used by Trudinger et al. (1997) to keep the model stable is no longer required, (vi) an exponentially-decreasing eddy diffusion following Severinghaus et al. (2001) has been added as an alternative to a well-mixed layer to model convective mixing near the surface, and (vii) the option to include dispersive mixing in the lock-in zone following Severinghaus et al. (2010) and Buizert et al. (2012).

Like the old version of the CSIRO firn model, the new version uses a reference frame that moves downwards (relative to the surface) with the ice. In the Supplement, we describe derivation of the firn model equations in the moving reference frame. While the derivation of the equations in the moving coordinates may seem complicated because we take care to distinguish between quantities and derivatives in fixed and moving coordinates, the final equations that are solved in the moving coordinate system are no more complicated than those in the fixed reference frame.

The main advantage of using moving coordinates is that advection in open firn and trapped bubbles is treated in a consistent way over all depths. This makes it easy to model any overlap of the bubble trapping with diffusive mixing (which can cause additional age spread of trapped air compared to firn air), as well as any variations in the ice properties that move with the ice, such as a melt layer (Trudinger et al., 1997), or seasonal variations in ice properties (Trudinger, 2000).

The version of the CSIRO model that participated in the NEEM firn model intercomparison (Buizert et al., 2012) did not include the upward flux of air due to compression. Including this flux significantly improves the fit to the NEEM data. All results in the present paper include this flux.

Forward integration of the model equations requires specification of

- The atmospheric concentration history, $c_X(t)$, (used as the upper boundary condition).
- Site information: temperature, T , atmospheric pressure, P , accumulation rate, A .
- Density profile, $\rho(z)$ (from which other quantities such as total porosity $s(z)$ and vertical ice velocity $v(z)$ can be derived; note that $v(z)$ also depends on A).
- Open porosity profile, $f(z)$ (we usually specify closed porosity $b(z)$ then $f(z) = s(z) - b(z)$).
- Molecular diffusivity profile of CO_2 , $D_{\text{CO}_2}(z)$.

- Diffusion coefficient of tracer X relative to the diffusion coefficient of CO_2 : $\gamma_X \equiv D_X/D_{\text{CO}_2}$. Multiplying this by the diffusivity profile of CO_2 gives the diffusivity profile of tracer X , $D_X(z)$.
- Profile of eddy diffusion, $D_{\text{eddy}}(z)$ (the same for all tracers, used for convective mixing near the surface or dispersive mixing in the lock-in zone, see next Section).
- Information about any other processes affecting mixing at a site, such as the depth of a well-mixed layer (representing convective mixing near the surface) or the date of creation and reduction of mixing of a melt layer.

The analysis in this paper involves solving the inverse problem of deducing $D_{\text{CO}_2}(z)$ from observations of multiple tracers. For some sites we also estimate parameters describing convective mixing or a melt layer. Additional cases described in the Supplement involve the estimation of γ_X or dispersion in the lock-in zone. While we refer to the calibration observations as concentrations throughout the paper, they are actually mole fraction in dry air for most tracers, and isotopic ratios for the others.

2.2 Method for estimating diffusivity

For the diffusive part of the firn column, the diffusivity is taken as the inverse tortuosity multiplied by the free-air diffusivity of CO_2 at the annual mean temperature and pressure of the site (Schwander et al., 1993). Tortuosity is a dimensionless parameter, and can be thought of as the distance travelled by the average molecule divided by the distance it would travel in free air. The inverse tortuosity is a site-specific and depth dependent property of the porous medium assumed to be independent of gas type, temperature and pressure, and is included in our CO_2 molecular diffusivity profile, $D_{\text{CO}_2}(z)$.

In the convective zone, diffusivity is greatly enhanced via macroscopic stirring. This stirring can be represented in a one-dimensional model by an “eddy” diffusivity, a concept from physical oceanography (Severinghaus et al., 2010). It has also been suggested recently that dispersion may occur in the deep firn, and this can also be modelled using eddy diffusivity (Severinghaus et al., 2010; Buizert et al., 2012). (Note that the term “eddy” should not be interpreted literally; the low Reynolds numbers in deep firn do not allow for turbulent air flow.) Eddy diffusivity affects all gases similarly (because macroscopic air movement advects all gases together in the viscous flow regime), while molecular diffusion through the tortuous pore space does not. Therefore, for studies involving comparisons of different gases in particular, the diffusivity is generally modelled as the sum of separately parameterised diffusivity related to tortuosity (which we refer to as molecular diffusivity or simply diffusivity) and diffusivity related to macroscopic stirring (eddy diffusivity). Molecular diffusivity is the dominant diffusion process at all firn sites, so much of our focus here will be on estimating the molecular diffusivity profile of CO_2 in firn in the least prescriptive way that we

can find, in order to fully explore the equifinality associated with molecular diffusivity. We also model other processes related to diffusivity when they are required to match observations, including eddy diffusivity, but these other processes have been specified in more prescriptive terms.

We specify (molecular) diffusivity as a function of open porosity. We assume that diffusivity varies monotonically with open porosity. As we use monotonic relationships for density versus depth and open porosity versus density, diffusivity also varies monotonically with depth. The main reason for choosing monotonic variation of diffusivity is that we need some kind of smoothness constraint for our automatic method of diffusivity estimation, to exclude the many possible solutions that reproduce the firn data well but are characterised by large and unrealistic oscillations in diffusivity (Rommelaere et al., 1997). The assumption of monotonic diffusivity is consistent with the assumption that the ice properties (density, open porosity and diffusivity) are time invariant. We are therefore considering the mean profile of the ice properties, and not seasonal or other short-term variations in firn microstructure which are not fully understood (e.g. the effect of impurities on densification, Hörhold et al. (2012)) and would be very difficult to specify or constrain. An important exception to the monotonic diffusivity assumption is a melt layer, such as that found at DE08-2 (Trudinger et al., 1997). We model a melt layer explicitly as a layer boundary (therefore moving with the ice) with reduced diffusivity, as described in the Supplement.

Other than the monotonic assumption, we wish to avoid being prescriptive about the form of the diffusivity function. For this reason we have chosen to define the diffusivity versus open porosity profile by interpolating with monotonic cubic splines (Wolberg and Alf, 2002) between about a dozen discrete points. The points are defined by fixing a diffusivity value and tuning the corresponding open porosity at which that diffusivity occurs (see Fig. 1). The exception is surface diffusivity, where we estimate the diffusivity corresponding to surface open porosity. The reason for tuning open porosity for fixed diffusivity, instead of tuning diffusivity for fixed open porosity, is to allow increased resolution as the diffusivity approaches zero (we do not know a priori for what values of open porosity the diffusivity will approach zero). Figure 1 shows the prior range for the open porosity values, f_n , corresponding to diffusivity $D_n = [0.0, 0.03, 0.05, 0.07, 0.1, 0.2, 0.5, 3.0, 20.0, 50.0, 110.0, 200.0, 300.0, 400.0] \text{ m}^2 \text{ yr}^{-1}$ and the surface diffusivity value. This range is used for the initial calculations at NEEM, and a reduced range is used for later calculations.

We use a genetic algorithm (Haupt and Haupt, 1998) to search for points that give the best agreement between modelled and measured firn concentrations of the reference tracers, by minimising the cost function, Φ , defined as the weighted root mean square mismatch

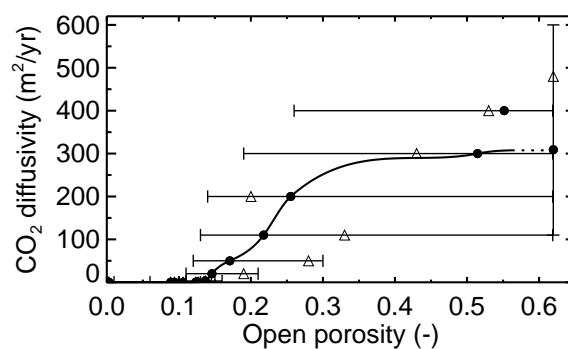


Fig. 1. The error bars show prior ranges for the porosity values (horizontal) and surface diffusivity (vertical) estimated with the GA. The solid line shows a monotonic smoothing spline fitted to the filled circles (ignoring the point at $400 \text{ m}^2 \text{ yr}^{-1}$ because it is greater than the surface diffusivity, see text). The dashed line shows diffusivity over the well-mixed layer (not used). The open triangles show a case with points that do not increase monotonically (so would not be tested in the firn model).

$$\Phi = \left(\frac{1}{N} \sum_{i=1}^N \frac{(m_i - d_i)^2}{\sigma_i^2} \right)^{\frac{1}{2}} \quad (1)$$

where N is the number of firn data, m_i and d_i are the modelled and measured firn concentrations, respectively, and σ_i are the one standard deviation uncertainties of the firn data (augmented to include other uncertainties as discussed in the next section).

Because we are only interested in monotonic solutions, and the interpolation by monotonic cubic splines will only work if the points are monotonic, any cases that the GA generates with points that are not monotonic (such as the case shown by the open triangles in Fig. 1) are not tested in the firn model, but instead assigned a very large cost function value that increases as points depart further from monotonic (to help guide the GA toward monotonic solutions early in the calculation). When fitting the monotonic spline (and also when checking for monotonicity of points), we ignore any points with diffusivity greater than the current surface value (e.g. in the case shown in Fig. 1 by the circles, the point at diffusivity = $400 \text{ m}^2 \text{ yr}^{-1}$ would be ignored). The prior ranges were chosen by trial and error to be large enough to include any solution that gives a good fit to the data.

The parameters we estimate are the open porosity values corresponding to the fixed diffusivity values, surface diffusivity, and any other diffusivity parameters required for the site, such as well-mixed layer depth, parameters defining the eddy diffusion profile, or the reduction of diffusivity associated with a melt layer that moves with the ice. For the GA calculation, we start with an ensemble of 500 randomly selected parameter sets, and run the calculation for 30 000 generations, with the GA breeding (i.e. combining) and mutating

(altering) parameter sets in the ensemble to obtain a new ensemble for each subsequent generation, analogous to genetics. Typically less than 10 % (and sometimes as low as 1 %, depending on the number of parameters being tuned and the ranges chosen) of the 500×30000 parameter sets tested will require a run of the firm model, the rest have non-monotonic points. It should be noted that our method is not very efficient at finding the one diffusivity profile that gives the best fit to the observations. Gradient-descent methods are more suited for that purpose. The GA, on the other hand, is well suited to locating a large range of diffusivity profiles and other parameters that fit the observed concentrations up to a chosen value of Φ .

To capture the range due to equifinality, as the GA runs we save the parameters and calculated depth profiles for any solution with Φ less than a chosen threshold. We can relate our choice of threshold for Φ to a confidence interval using the F ratio test (Árnadóttir and Segall, 1994; Seber and Wild, 2005). The value of Φ that corresponds to a $100 \times (1 - \alpha)$ % confidence level is determined from

$$\Phi^2 = \Phi_{\text{opt}}^2 \left[1 + \frac{p}{n-p} F(p, n-p, 1-\alpha) \right] \quad (2)$$

where Φ_{opt} is the optimal value of Φ , p is the number of parameters estimated, n is the number of observations and $F(p, n-p, 1-\alpha)$ is the F distribution with p and $n-p$ degrees of freedom at a $100 \times (1-\alpha)$ % confidence level (Draper and Smith, 1981). Therefore all models with Φ less than this threshold are consistent with the optimal model at the specified confidence level. Note that these confidence intervals are only approximate, as they assume Gaussian errors and linear variation of Φ around the minimum. They are also subject to our assumption of monotonic diffusivity profiles, and to our choice of processes to represent which determines the number of parameters we estimate. We use the F ratio test to guide our choice of Φ for all cases we consider.

For most of our calculations we consider solutions with Φ up to the 68% confidence level (i.e. ± 1 standard deviation), and refer to these as the accepted solutions. The threshold for Φ is a somewhat subjective choice, and various other options would be equally valid. For NEEM observations, our lowest value of Φ is 0.74, and the 68 % confidence level corresponds to $\Phi = 0.78$. We could also have chosen thresholds such as 0.92 (approximately the range covered by all models in Buizert et al. (2012)) or 0.8 (the range covered by four of the six models in Buizert et al. (2012), where the two models that had $\Phi > 0.8$ were the old version of the CSIRO model that excluded the upward flux of air due to compression, and the OSU model for which Buizert et al. (2012) attributed the higher Φ to fewer degrees of freedom in the tuning procedure). When selecting a threshold, we need to keep in mind the fact that our model is not a perfect representation of the real firm processes, and the data inevitably contain errors, so we should avoid retaining only the very best solutions. In light of this, the choice of 68 % confidence level is probably

too low to reflect the full uncertainty. However, we made this choice because it includes only solutions that are a very good fit to the observations yet we still see an interesting amount of spread due to equifinality. We show the sensitivity of our results to the choice of Φ threshold in a couple of cases in Sects. 3.1 and 3.2.

To look at how well different combinations of tracers constrain diffusivity we will tune diffusivity using subsets of the available tracers for a site (with the set of all tracers being one of the subsets). We determine the threshold for Φ using all tracers and then use the same level for the other subsets (i.e. for NEEM we would use 0.78 for all subsets of tracers, even though some subsets have a best value of Φ below 0.74). By applying the same threshold to different subsets with different numbers of observations, our results for each subset correspond to different confidence levels. For example, our initial NEEM calculations have 16 parameters (14 open porosity values for specified diffusivity levels, the surface diffusivity value and well-mixed layer depth) and 220 observations (ten tracers), the threshold that corresponds to 68 % confidence level for all observations has a very low confidence level with 44 observations (two tracers), implying that a significantly higher threshold would be needed to maintain the 68 % confidence level. However, we choose to use a fixed threshold so that we can compare the results of calibrations that fit the observations to the same degree. The alternative of changing the threshold for different subsets to correspond to the same confidence level is equally valid, but in that case a significant amount of the difference in spread between subsets will be due to the different number of observations, which is not what we are interested in at present.

To understand the range of solutions generated, it is useful to consider how solutions can vary within the range up to the Φ threshold. There is a tendency for most variation to occur in tracers (or depth ranges) that have unique information and less variation in tracers (or depth ranges) that contain redundant information. This point is the key to understanding a number of features of the results.

When choosing which tracers to use for calibration, we note that the ability of a reference tracer to constrain the diffusivity profile depends on two things. The first is the amount of variation with depth the tracer has in the firm, which is largely a consequence of the particular atmospheric history of the tracer, but also depends on the processes in the firm that we wish to learn about. This is to do with the signal and how much information on firm processes is contained in the depth profiles of each of the reference tracers. The second is the total uncertainty in the data, based on analytical precision, error in atmospheric reconstructions and the ability of the model to capture processes important for that tracer. This is about the noise and how much this obscures the signal. Analysis of both the signal and the noise are important for determining which tracers are most useful, as is the case with any inverse study. Tracers are expected to differ from site to site in their relative importance for constraining diffusivity,

depending on the relative timescales of atmospheric change and firn closure, and the relative uncertainties in the calibration and forcing data.

In our synthetic calculations, as well as testing and learning about our method, we wish to focus on the first aspect, that is, how useful different tracers are for calibrating diffusivity due to their different atmospheric histories. This motivates our choice of errors for the synthetic data, which we set to have similar characteristics for all tracers, as described in the next section. Clearly, cases with real data must also consider the fact that some tracers can be measured more accurately than others, and some have greater uncertainty in their atmospheric history. However, by looking first at the effect of atmospheric history without the complications of different errors for different tracers, we can learn more about the principles for selecting the most useful tracers.

2.3 Observations and model inputs

2.3.1 NEEM

Most of our calculations relate to the NEEM 2008 firn air campaign (Buizert et al., 2012). From that study we use firn air data for 10 tracers from the S2 borehole at NEEM in Greenland (CO_2 , CH_4 , SF_6 , CFC-11, CFC-12, CFC-113, CH_3CCl_3 (methyl chloroform), HFC-134a, $^{14}\text{CO}_2$ and $\delta^{15}\text{N}_2$), uncertainty estimates (based on a range of contributions including analytical precision and uncertainty in atmospheric histories), and the physical firn characteristics (see Table 1). The fit to the reference tracers as calculated with Eq. (1) using the NEEM observations and uncertainties given in Buizert et al. (2012) is denoted Φ_{N} .

The concentration histories for the gases considered in this paper are described and shown in the Supplement. We use the relative diffusion coefficients from Matsunaga et al. (1993, 1998, 2002). For all of the sites we consider, we follow Buizert et al. (2012) by converting firn measurements and atmospheric histories of $\Delta^{14}\text{CO}_2$ in permil to $^{14}\text{CO}_2$ in ppm, as well as modelling $\delta^{15}\text{N}_2$ using a single $^{15}\text{N}^{14}\text{N}$ tracer.

In order to test our method for estimating diffusivity, and to see how well different tracers constrain diffusivity in an ideal situation without the complications of real observations and an imperfect model, we created synthetic data using the NEEM physical firn characteristics and the tracers that were measured at NEEM. We generated what we call “true concentrations” by running our firn model with the molecular diffusivity tuned to 2008 S2 NEEM observations with the LGGE-GIPSA firn model in Buizert et al. (2012), and a 3.66 m well-mixed layer. Due to differences in firn models, the same diffusivity profile used in two different firn models is expected to give different results, so we are not trying to recreate the NEEM concentrations here, however we do expect the results to be similar to NEEM. The LGGE-GIPSA diffusivity from Buizert et al. (2012) was chosen because, unlike most other models in that study, LGGE-GIPSA did

not include dispersion in the lock-in zone, so we could create NEEM-like concentrations using only molecular diffusion, and it gave us an independent diffusivity profile, which we refer to as the “true diffusivity”, to try to recover with our method.

We created two datasets of synthetic observations, and from these we wish to recover the “true” depth profile of diffusivity as well as the “true” depth of the well-mixed layer (i.e. 3.66 m). The first dataset, denoted “Synthetic A”, has 22 observations for each of the 10 tracers, at depths matching the NEEM observation depths excluding the surface. We added uncorrelated random noise to our true concentrations sampled at the measurement depths. The random noise was shifted and scaled for each tracer so that the mean of the noise added to each tracer was zero and the standard deviation was exactly 0.5 % of the range of each tracer (maximum minus minimum values of the true concentrations at the observation depths). The comparison of modelled concentration profiles with the Synthetic A observations using Synthetic A uncertainties is denoted Φ_{A} . The data uncertainty σ_i used in Φ_{A} matches the standard deviation of the noise added to each tracer. This means that the true solution corresponds to exactly $\Phi_{\text{A}} = 1.0$ for all tracers separately and together (due to the noise).

The second set of synthetic observations, denoted “Synthetic B”, has the same true concentration as Synthetic A, but observation depths that correspond exactly to the types and measurement depths for the S2 borehole in the NEEM intercomparison (between 15 and 23 measurements of each tracer), and error that consists of random error plus a systematic offset added to each tracer plus a systematic error that increases linearly with depth from zero at the surface (see Supplement for details). Depending on the random values generated, some tracers will have larger systematic errors added than others. Synthetic B is more like reality, where systematic errors probably dominate, than Synthetic A. This case is similar to the real NEEM case in terms of the general error characteristics (but not in terms of the specific error characteristics for each tracer). The comparison of modelled concentration profiles with the Synthetic B observations using Synthetic B uncertainties is denoted Φ_{B} . While most of our synthetic calculations will use the Synthetic A observations, allowing us to explore which tracers are most useful based on their atmospheric histories, we include a couple of calculations with the Synthetic B observations to see what difference larger, systematic errors make to the results.

For the NEEM calculations (with both synthetic and real observations) we look at the diffusivity estimated with observations of all ten tracers, as well as other subsets of tracers as listed in Table 2. The subsets were chosen by starting with a minimum set that might be used, CH_4 and $\delta^{15}\text{N}_2$; these are two tracers that have been measured at almost all (if not all) firn air sites. We then add one tracer at a time with the aim that each new tracer will add new information to improve the constraint on diffusivity. This is based on the expectation

Table 1. Site characteristics.

Site	Date sampled	Accum. rate kg m ⁻² yr ⁻¹	Temp. °C	Press. hPa	LID m
NEEM	Jul 2008	198.8	-28.9	745	63
DE08-2	Jan 1993	1100	-19	850	73
DSSW20K	Jan 1998	150	-20.7	850	43
South Pole	Jan 1995, 2011	74	-51	681.5	114

Table 2. Subsets of tracers used for calibration of diffusivity at NEEM (pseudo and real observations), where the name reflects the number of tracers used.

Subset	Tracers
Two	$\delta^{15}\text{N}_2$, CH_4
Three	$\delta^{15}\text{N}_2$, CH_4 , SF_6
Four	$\delta^{15}\text{N}_2$, CH_4 , SF_6 , HFC-134a
Five	$\delta^{15}\text{N}_2$, CH_4 , SF_6 , HFC-134a, CH_3CCl_3
Ten	$\delta^{15}\text{N}_2$, CH_4 , SF_6 , HFC-134a, CH_3CCl_3 , CO_2 , CFC-11, CFC-12, CFC-113, $^{14}\text{CO}_2$

that tracers with atmospheric records covering different periods will constrain diffusivity in different parts of the profile. CO_2 and $^{14}\text{CO}_2$ were excluded from the subsets because of concerns of possible contamination in Northern Hemisphere firn and in situ production of ^{14}C by cosmic rays, even though this is not an issue with synthetic data. For Synthetic A, we also compare results using each of the ten tracers separately.

In comparing these subsets, we will compare both the best case for each subset as well as the range of accepted solutions. While they often lead to broadly similar conclusions, these two ways of looking at the results are affected by different things, and are therefore both useful.

2.4 DE08-2

DE08-2 is located on the east side of Law Dome, East Antarctica (Etheridge et al., 1996, 1998; Trudinger et al., 1997) and firn air was sampled in 1993. Site parameters are given in Table 1. For density versus depth we use a spline fit to measurements from Etheridge and Wookey (1989) and Etheridge et al. (1996). For closed porosity versus density we use a spline fit to measurements by J.-M. Barnola (unpublished data) with a cut bubble correction (J.-M. Barnola, personal communication, 1999) which gives a rapid reduction of open porosity to zero around 90 m, in line with the observation that there was still good flow of firn air at 85 m, but no air could be pumped from the firn at 90 m (Etheridge et al., 1996).

The ice structure at DE08-2 shows a melt layer at 8.7 m below the surface. Trudinger et al. (1997) found that the agreement between modelled and measured tracers at DE08-

2 (SF_6 in particular) was significantly improved by including a melt layer that originated at the surface in the 1989–1990 summer and moved with the ice with a reduction in the diffusive flux of about 80 %. Here we model the melt layer as originating at the surface in 1989.77 and moving with the ice as described in the Supplement. We tune the reduction in mixing along with the diffusivity profile using measurements of CO_2 , CH_4 , SF_6 , $\delta^{15}\text{N}_2$ and $^{14}\text{CO}_2$ (Etheridge et al., 1996; Trudinger et al., 1997; Levchenko et al., 1997). DE08-2 firn measurements and the uncertainties we use are given in the Supplement. The fit to these measurements is denoted Φ_{D} .

2.5 DSSW20K

DSSW20K is located on the lower-accumulation-rate west side of Law Dome, East Antarctica and firn air was sampled in January 1998 (Smith et al., 2000; Sturrock et al., 2002; Trudinger et al., 2002). The firn column at DSSW20K is relatively short, with a lock-in depth of about 43 m. We use a spline fit to DSSW20K density measurements, and the DE08-2 closed porosity spline. DSSW20K has a region of convective mixing near the surface that was modelled by Trudinger et al. (2002) using a 4 m well-mixed region. Here we use either an exponentially-decreasing eddy diffusion or well-mixed layer and tune the relevant parameters along with the diffusivity profile with measurements of CO_2 , CH_4 , SF_6 , $\delta^{15}\text{N}_2$, $^{14}\text{CO}_2$, HFC-134a, HCFC-141b, CFC-11, CFC-12, CFC-113 and CH_3CCl_3 (Smith et al., 2000; Sturrock et al., 2002; Trudinger et al., 2002). The DSSW20K firn measurements and the uncertainties we use are given in the Supplement. The fit to these measurements is denoted Φ_{W} .

Southern Hemisphere atmospheric histories of SF_6 , CH_3CCl_3 and the CFCs prior to 1978 (when the Cape Grim Air Archive and/or Cape Grim in situ measurements began) are based on emissions estimates rather than atmospheric measurements (see Supplement). For DSSW20K, we choose not to use concentration measurements of these tracers below 45 m in our model calibration because of the greater uncertainty in their atmospheric histories. The alternative approach is to include the firn observations with larger uncertainties to reflect the greater uncertainty in the atmospheric histories – this is the approach we have used for NEEM, because we use exactly the same dataset as Buizert et al. (2012) to allow direct comparison, and this is their approach. We will test

Table 3. Weighted RMS mismatch of modelled concentrations from Synthetic A observations and from the true concentrations for each tracer for each of the five Subsets in Table 2. Modelled concentrations are for the case with lowest Φ for each experiment. Numbers in bold are for tracers that were fitted for that particular Subset. As well as RMS mismatch for each tracer, we also show the RMS mismatch for all ten tracers together (All), and the RMS mismatch for only those tracers that were fitted in that experiment (Fitted). The true solution corresponds to the $\Phi_A = 1.0$ and $\Phi_{At} = 0.0$.

	Mismatch from synthetic observations, Φ_A					Mismatch from true concentrations, Φ_{At}				
	Two	Three	Four	Five	Ten	Two	Three	Four	Five	Ten
CO ₂	2.98	1.36	1.04	1.00	1.06	2.74	0.67	0.33	0.38	0.36
CH ₄	0.77	0.93	0.92	1.04	0.93	0.74	0.61	0.53	0.73	0.50
SF ₆	5.55	1.04	0.82	0.87	1.05	5.15	1.11	0.32	0.29	0.37
CFC-11	2.16	1.17	1.00	1.03	0.98	1.98	0.70	0.31	0.31	0.28
CFC-12	2.23	1.30	1.12	1.09	1.14	1.88	0.68	0.31	0.32	0.30
CFC-113	3.64	1.30	1.08	1.16	1.07	3.21	0.92	0.28	0.32	0.33
HFC-134a	8.44	2.05	1.03	0.99	1.24	8.41	1.59	0.41	0.33	0.51
CH ₃ CCl ₃	9.23	2.41	1.25	1.04	1.09	9.20	2.19	0.67	0.50	0.70
¹⁴ CO ₂	1.17	1.09	1.05	1.11	0.73	0.80	0.28	0.45	0.74	0.63
$\delta^{15}\text{N}_2$	0.93	0.95	0.97	0.97	0.99	0.23	0.11	0.11	0.12	0.10
All	4.71	1.44	1.03	1.03	1.03	4.57	1.06	0.40	0.45	0.44
Fitted	0.86	0.97	0.93	0.99	1.03	0.49	0.61	0.34	0.39	0.44

the sensitivity of calibration results to including or excluding these observations for NEEM.

2.6 South Pole

Firn air was collected from South Pole, Antarctica in 1995 (Battle et al., 1996; Butler et al., 1999) and again in 2001 (Butler et al., 2001; Aydin et al., 2004; Sowers et al., 2005). Following previous work for South Pole, we treat these as two separate sites. We use the site properties in Table 1, a spline fit to South Pole density measurements (different for 1995 and 2001 drillings) from Mark Battle (personal communication, 2012) and the DE08-2 closed porosity spline. For South Pole 1995 we use measurements of CO₂, CH₄, SF₆, $\delta^{15}\text{N}_2$, CFC-11, CFC-12 and CH₃CCl₃ (Battle et al., 1996; Butler et al., 1999) and the fit is denoted Φ_{S95} . For South Pole 2001 we use CO₂, CH₄, SF₆ and $\delta^{15}\text{N}_2$ (Aydin et al. (2004); Sowers et al. (2005); Witrant et al. (2012); M. Battle, personal communication, 2012), and the fit is denoted Φ_{S01} . For both South Pole sites we omit firn concentration measurements corresponding to ages before 1978 for tracers with atmospheric histories based on emissions estimates.

3 Results

3.1 Synthetic calculations

We first consider the case with synthetic observations for NEEM and only uncorrelated random noise (Synthetic A). Figure 2 shows the concentration profiles for Synthetic A using only observations of CH₄ and $\delta^{15}\text{N}_2$ for calibration (Subset Two, in panels a–j), and using all ten tracers for calibration (Subset Ten, in panels k–t). Figure 3 shows the diffusiv-

ity profiles and CH₃CCl₃ for all five Subsets in Table 2. We show results for the threshold $\Phi_A = 1.08$ corresponding to the 68% confidence level using observations of all ten tracers. We also include the range for the alternative threshold $\Phi_A < 1.25$ to show the sensitivity of the range to the choice of threshold.

If all of the accepted solutions were plotted, the region between the black dotted lines would be almost completely shaded. We show 19 representative solutions covering different values of Φ_A over the range between the lowest value (0.86 for Two and 1.03 for Ten) and 1.08, and include solutions that differ most from the best solution and each other. Representative solutions such as these can be used to represent equifinality in the firn model when it is used to infer atmospheric histories of other trace gases.

Table 3 gives the weighted RMS mismatch of each best solution from the noisy synthetic observations for each tracer separately and together for the five Subsets considered, with numbers shown in bold if that tracer was used for calibration in that experiment. Table 3 also gives the weighted RMS mismatch of each best solution from the true concentrations weighted by the same data uncertainties, denoted Φ_{At} .

The diffusivity profile above 65 m was not well constrained with CH₄ and $\delta^{15}\text{N}_2$ alone. CH₄ does not vary much with depth above 65 m, so there is not much information to distinguish between diffusion in different parts of the firn above 65 m. $\delta^{15}\text{N}_2$ is quite sensitive to the depth of the well-mixed layer, but it is fairly insensitive to the diffusivity profile within the parameter range considered. Diffusion around 65–70 m is quite tightly constrained (small spread of solutions) by CH₄ and $\delta^{15}\text{N}_2$. Calibrating with only these two tracers, the best estimate of diffusivity and the best concentration profiles of most tracers not used in the GA are quite

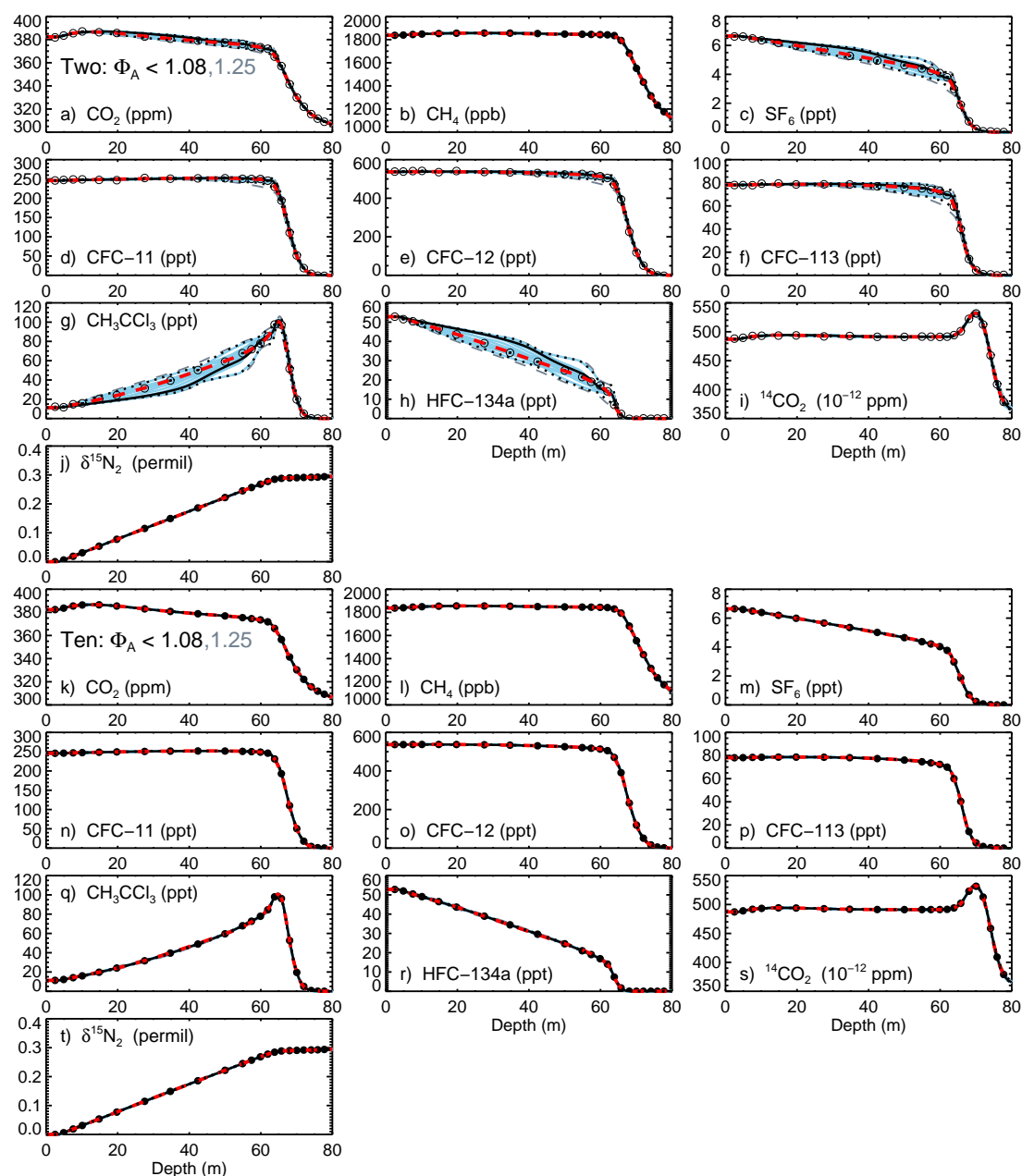


Fig. 2. Modelled concentrations for all ten tracers, where diffusivity was calibrated using Synthetic A observations of only CH_4 and $\delta^{15}\text{N}_2$ (Subset Two) (a–j), and using all ten tracers (Subset Ten) (k–t). The black solid line is the solution with the lowest Φ_A , black dotted lines show upper and lower ranges of all accepted solutions ($\Phi_A < 1.08$), the blue lines are 19 representative solutions and the red dashed line shows the true solution. The grey dashed lines show the upper and lower ranges of all solutions with $\Phi_A < 1.25$. Observations used to tune diffusivity are shown with filled circles, observations not used in the GA are shown by open circles. Error bars for the observations (1σ) are plotted but are too small to see relative to the symbols.

different from the true profiles ($\Phi_{A_t} = 4.57$). There is also significant spread in solutions with $\Phi_A < 1.08$, with spread being greatest for CH_3CCl_3 that has a rapid recent atmospheric decrease, and HFC-134a that has a rapid recent rise in the atmosphere. $^{14}\text{CO}_2$ has a small spread above the lock-in depth – with little variation with depth above 65 m, this part

of the $^{14}\text{CO}_2$ profile would provide little additional information if used for calibration. However, the peak in $^{14}\text{CO}_2$ at NEEM is below the lock-in depth, and there is some spread in solutions around the peak and even more spread around 75–80 m, suggesting that it could be a useful tracer in this region at NEEM, particularly as the other tracers have either

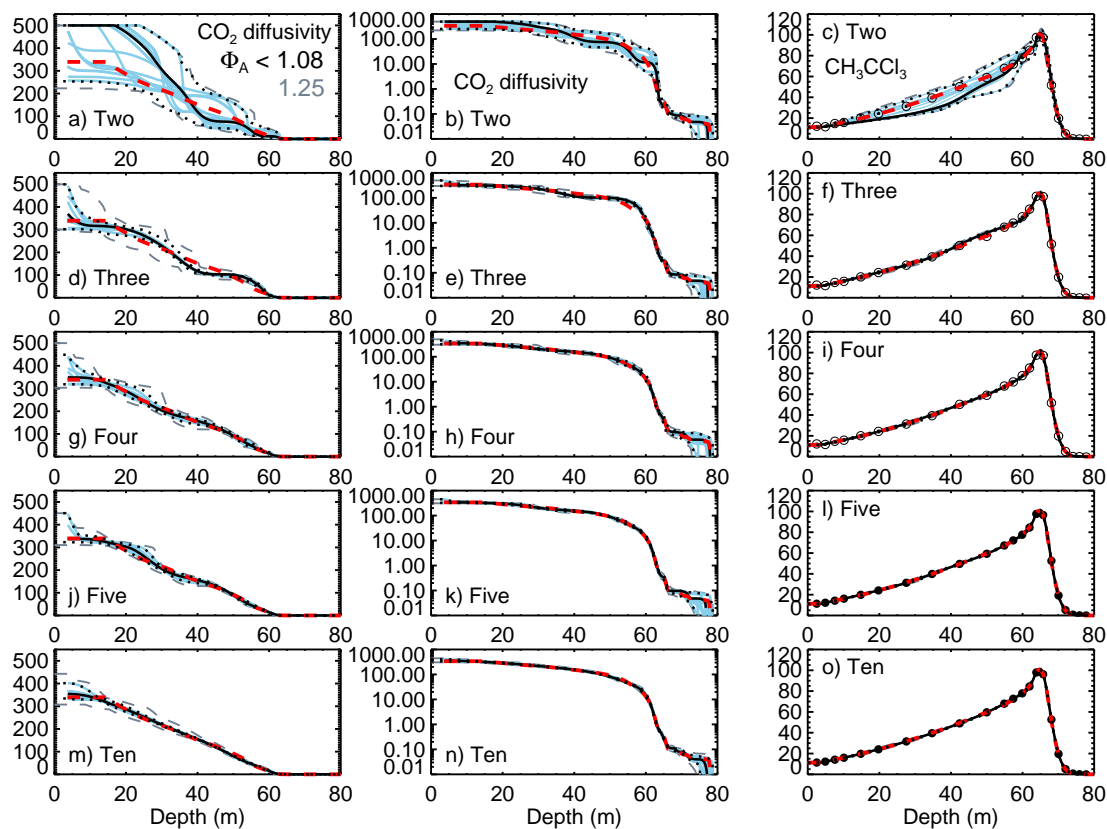


Fig. 3. CO₂ diffusivity profile estimates in m² yr⁻¹ (left column on a linear scale and middle column on a log scale) and calculated CH₃CCl₃ in ppt (right column) from the GA using Synthetic A observations for all five Subsets in Table 2. Line styles are as in Fig. 2 and results are for $\Phi_A < 1.08$, with the grey dashed lines showing the range for $\Phi_A < 1.25$.

zero concentration or very little spread there. There is only a small change in spread between the thresholds for Φ_A of 1.08 (black dotted lines) and 1.25 (grey dashed lines), and we would reach very similar conclusions with either threshold.

When all ten tracers were used for calibration, the diffusivity and concentration profiles for the best solution were close to the true profiles ($\Phi_{Ar} = 0.44$) and the weighted RMS mismatch from the noisy observations (Φ_A) was around 1.0. The spread in concentrations of these ten tracers for solutions with Φ_A up to even 1.25 was very small.

There is a clear difference between Subset Two and the other four subsets in Fig. 3. Adding just SF₆ to the calibration (Subset Three) makes a large difference in reducing the spread of accepted solutions, and in bringing the best solution closer to the true profile. There is further improvement to the best solution by adding HFC-134a (Subset Four), but little further improvement beyond this up to the ten tracers, other than some reduction in spread below 70 m from Subset Five to Ten (we will see later that this is mostly due to adding ¹⁴CO₂). It is encouraging that in this simplest case with uncorrelated errors, we can reconstruct diffusivity well

with four or more tracers and our specification of diffusivity using cubic splines between a small number of points.

Recall that the true diffusivity corresponds to Φ_A of exactly 1.0 for all tracers separately and together (due to the noise). Values of Φ_A for the best solutions of the different Subsets (for observations actually fitted) vary from 0.86 to 1.03, generally lower than 1.0 for two to four tracers and around 1.0 for five and ten tracers. The lower values of Φ_A are achieved by fitting to some of the noise, i.e. overfitting the noisy data. Diffusion in the firn leads to smooth profiles of concentration with depth, still the GA does manage to find solutions that are closer to the noisy observations, quantified by the RMS difference, than the true solution. Overfitting decreases as more tracers are added, and by five tracers we have essentially stopped overfitting the noisy data.

The depth of the well-mixed layer was estimated as part of the inversion. When $\delta^{15}\text{N}_2$ was used for calibration the estimated well-mixed layer depth for the best solution was always close to the true value of 3.66 m with a clear increase in Φ_A moving away from this value. In a calibration calculation without $\delta^{15}\text{N}_2$ but with the other nine tracers, the well-mixed layer was not quite as well resolved.

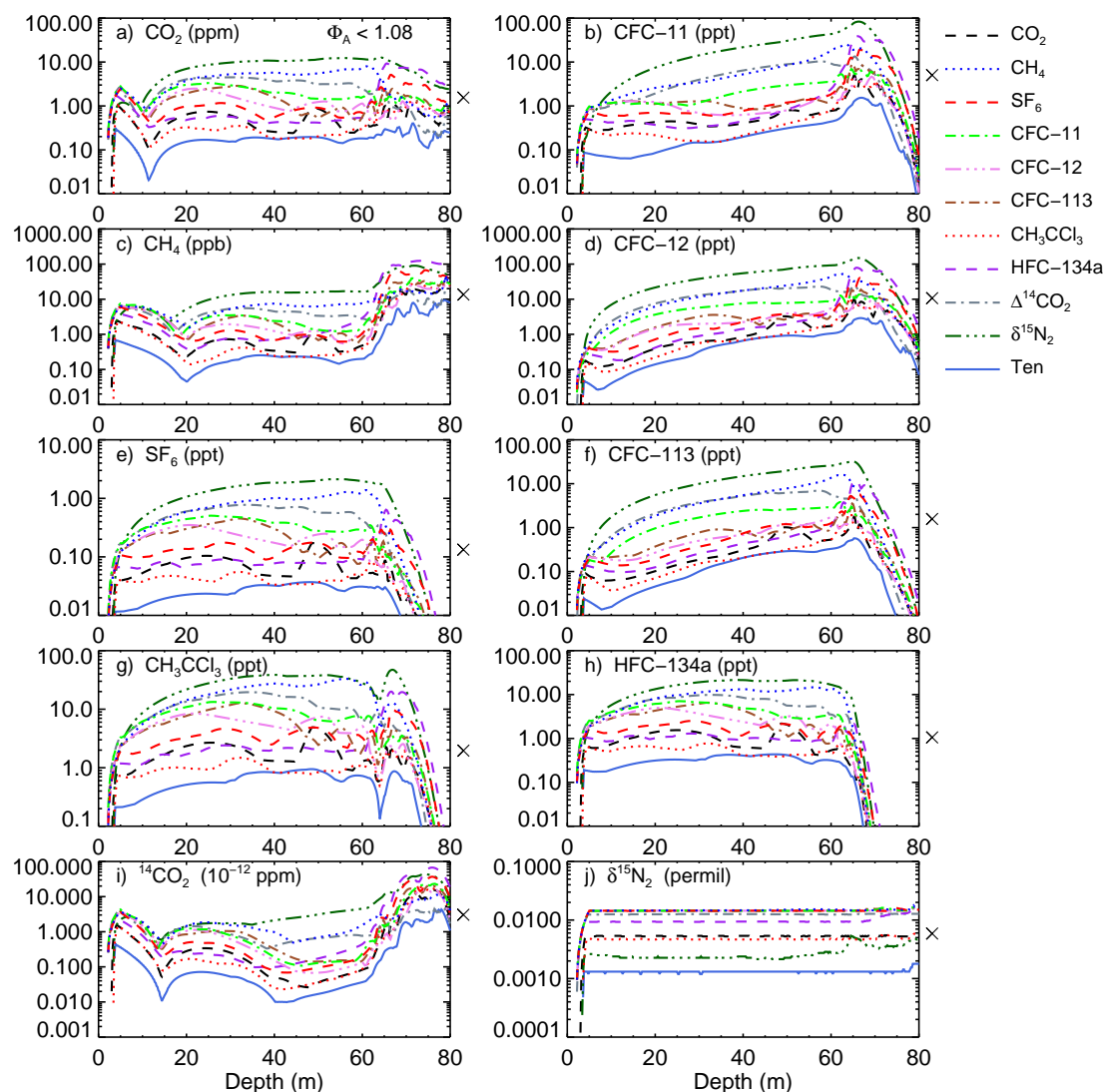


Fig. 4. Range of concentration for all 10 tracers, for all accepted solutions ($\Phi_A < 1.08$) with Synthetic A observations, with each tracer used for calibration on its own and Subset Ten described earlier. Each panel shows the range calculated for one tracer, and each line in that panel corresponds to a different tracer (or Subset Ten) used for calibration. To the right of each panel, the cross symbol shows 2σ (twice the standard deviation of the error applied to the observations of that tracer).

We also tested calibration of diffusivity and the well-mixed layer depth with each of the ten tracers taken separately. Figure 4 shows the range (i.e. maximum minus minimum) of accepted concentration profiles ($\Phi_A < 1.08$) for all ten tracers when only one tracer was used for calibration (i.e. this corresponds to the difference between the black dotted lines in earlier plots). To the right of each panel, the cross symbol shows twice the standard deviation of the error applied to observations of each tracer (2σ), for comparison with the full range of accepted solutions.

By far the best constraint above 60 m was provided by CH_3CCl_3 , as indicated by both the small spread in the other tracers and the good fit of the best solution to all tracers ($\Phi_A = 1.59$ and $\Phi_{At} = 1.32$, where a significant fraction of

the mismatch comes from not fitting $^{14}\text{CO}_2$, CO_2 and CH_4 in the lock-in zone, as well as $\delta^{15}\text{N}_2$). CO_2 provided a very good constraint throughout the whole firn. HFC-134a and SF_6 provided good constraints on the other tracers through most of the firn (to around 62 m), but weak constraints below this where concentrations are small. With little variation with depth above 60 m the CFCs do not provide strong constraints for the tracers with rapid atmospheric growth rates. Constraining diffusivity with either CH_4 or $^{14}\text{CO}_2$ gave a large spread in the other tracers above 60 m, but small spread below this, and in fact calibrating with just $^{14}\text{CO}_2$ gave the smallest spread for all tracers below 70 m. On its own, $\delta^{15}\text{N}_2$ has the weakest constraint on diffusivity and therefore leads to the largest spread in the other nine tracers, and its best case

has a poor fit to the other tracers. In addition, $\delta^{15}\text{N}_2$ is poorly constrained by the other tracers. This is as expected, as $\delta^{15}\text{N}_2$ is not very sensitive to molecular diffusion, but is strongly affected by convective mixing near the surface, unlike the other tracers.

Comparison of the Synthetic A data error (shown by the crosses) with the spread of that tracer for each case indicates whether Synthetic A observations of that tracer are likely to be constraining, and leads to similar conclusions as those just discussed. For example, the range in $^{14}\text{CO}_2$ in the lock-in zone for most cases in Fig. 4i is larger than the data error for $^{14}\text{CO}_2$, suggesting that Synthetic A lock-in zone observations of $^{14}\text{CO}_2$ are useful. In contrast, the spread in $^{14}\text{CO}_2$ above 60 m is less than the data error, suggesting that $^{14}\text{CO}_2$ is not so useful there, as we have already seen.

In summary, CH_3CCl_3 , HFC-134a, SF_6 and CO_2 stand out in the Synthetic A calculations as important above the lock-in zone, and tracers that have been in the atmosphere longer like CH_4 , CO_2 and especially $^{14}\text{CO}_2$ with its variation due to the bomb pulse are important in the lock-in zone. These two groups of tracers contain complementary information, and at least one representative from each group would be needed to constrain diffusivity over the full range, with the best of each group being CH_3CCl_3 and $^{14}\text{CO}_2$ respectively. $\delta^{15}\text{N}_2$ is also useful to constrain the convective zone. Based on their atmospheric histories alone, we could identify a minimum subset of tracers to constrain NEEM diffusivity as CH_3CCl_3 , $^{14}\text{CO}_2$ and $\delta^{15}\text{N}_2$, although with only three tracers, overfitting the noise in the data may be a problem.

Witrant et al. (2012) calibrated their firn model with the same tracers taken one at a time but using real NEEM observations and uncertainties, as well as leaving one tracer out at a time from the full set. Like us, they noted the value of CH_4 and $^{14}\text{CO}_2$ in the lock-in zone, and the lack of constraint from CFC-11 and CFC-113 in the upper firn due to their small variation of concentration with depth. However, when we calibrated to synthetic CH_3CCl_3 our best solution gave a good fit to the other tracers above 60m, but the case in Witrant et al. (2012) calibrating to real CH_3CCl_3 did not give a good fit to tracers such as HFC-134a, SF_6 and the CFCs. In their case, calibrating with real HFC-134a data did not fit SF_6 well, unlike our synthetic case. The differences between the synthetic and real cases are expected to be due to greater errors or inconsistencies in the firn model, data or atmospheric histories in the real case, and more detailed comparison of a real and synthetic case, ideally with the same firn model, would be needed to understand the differences properly. Based on this comparison, our synthetic calculations show where we can make the most improvement to our calibrations – e.g. CH_3CCl_3 has the potential to provide a very strong constraint on firn diffusivity if we can reduce the errors and inconsistencies associated with real data.

The diffusivity profile is a model-dependent quantity and of secondary importance compared to the age distribution which characterizes the firn air transport, and is more useful

for comparison between firn models and use in inversions for atmospheric histories (Rommelaere et al., 1997; Trudinger et al., 2002). We quantify the width of the age distribution with the spectral width, Δ , defined as

$$\Delta^2(z) = \frac{1}{2} \int_0^{\infty} (t - \Gamma)^2 G(z, t) dt \quad (3)$$

where t is time, z is depth and Γ is the mean of the age distribution, G (Hall and Plumb, 1994; Trudinger et al., 2002). The spectral width for an asymmetric distribution is analogous to the standard deviation of a normal distribution.

Figure 5a and b show the variation with depth of the range in the spectral width of CO_2 age distributions, calculated using representative solutions obtained by calibrating to subsets of Synthetic A observations, for two different thresholds for Φ_A (1.08 and 1.25). The range is calculated from our ensemble of 20 representative solutions, which we use to represent the full range of diffusivity solutions (recall that they were selected to cover roughly the same diffusivity range as the full set of solutions). The cases shown are the five Subsets in Table 2 plus the case calibrating to CH_3CCl_3 alone, and the case with $\delta^{15}\text{N}_2$, CH_3CCl_3 and $^{14}\text{CO}_2$ (our minimum subset defined above). Note that although we consider only age distributions for CO_2 , they are not at all affected by the atmospheric history of CO_2 , instead they are only affected by the firn diffusion processes for CO_2 and therefore characterise the firn transport processes well for all tracers.

In Fig. 5a and b we see that going from Subset Two to Three greatly reduces the spread in the calculated age distributions above 60 m, with a small further improvement going from Subset Three to Four, but little reduction beyond this for most of the firn. Below 70 m, Subset Ten is significantly better than the other (original) Subsets, because it is the only subset that includes $^{14}\text{CO}_2$, the best parameter to constrain the diffusivity in that depth range. We also saw these features in the diffusivity profiles (Fig. 3). In addition, between 10 m and 65 m, CH_3CCl_3 alone constrains the age distribution roughly as well as all ten tracers with the Synthetic A observations, but below 65m the spread is large, and below 70m it is the worst of these cases. Combining the three tracers CH_3CCl_3 , $^{14}\text{CO}_2$ and $\delta^{15}\text{N}_2$ results in a very low spread of solutions throughout the firn, sometimes lower than for Subset Ten in the deep firn, presumably because these three tracers give more equal weight to tracers with information content above and below the lock-in zone than Subset Ten. (Subset Ten has many more tracers with information content mainly above 70 m, so with the same Φ threshold used for both cases, the GA can find solutions with Subset Ten that relax the fit to the couple of tracers with variation below 70 m and still have Φ_A below the threshold.) The features are generally similar for these two choices of threshold. Increasing the threshold tends to increase the spread most in regions/tracers with unique information.

It is very important to remember that the Synthetic A data has small, Gaussian errors that are the same for all tracers relative to their overall range of variation in the firn. These calculations tell us about the information content of the tracers due to their atmospheric histories largely without consideration of data errors that would be likely to be found in a realistic case. Nonetheless, they have been useful to allow us to explore our method and learn about the principles for selecting the best tracers based on their atmospheric history.

We also consider a couple of calculations with the Synthetic B case that has the same true diffusivity as Synthetic A but more complicated data errors. Synthetic B data and results are described in more detail in the Supplement. With a lowest value of Φ_B of 0.8 for Subset Ten, we consider solutions with Φ_B up to the 68 % confidence level of 0.84 as well as an alternative threshold of 1.0 for all Synthetic B subsets. Figure 5c shows the range in the spectral width of the age distribution with depth for solutions with $\Phi_B < 0.84$ in the original five Subsets.

Compared to Synthetic A, the results for Synthetic B are further from the true solution, and the spread in diffusivity for the range of accepted solutions is significantly greater (from comparing Fig. 3 with Fig. S3 in the Supplement). None of the Subsets using Synthetic B observations is closer to the true solution than Subset Three with Synthetic A observations and uncertainties. This is not surprising, because of the larger and more complicated (non-Gaussian) errors in the Synthetic B observations. Despite the results being worse than for Synthetic A, we do manage to reconstruct the diffusivity profile reasonably well for four or more tracers. Compared to Subset Five, Subset Ten has an improvement in the best solution in the deep firn as well as a reduction in the spread of accepted solutions in this region. Overall, the variation in the best solution and spread of solutions for Synthetic B subsets is quite similar to those for Synthetic A.

For Synthetic B, the true diffusivity gives Φ_B of 0.81, and the best cases have values of Φ_B for fitted observations between 0.62 and 0.8, with only Subset Ten having Φ_B near the true value of 0.81. In both Synthetic A and B, the model fits the data better than the true values for small numbers of tracers. Overfitting of the data is more of a problem for Synthetic B than A, as expected, indicated by the lower values of Φ relative to Φ for the true solution. We needed five tracers to avoid it with Synthetic A but at least six and possibly up to ten tracers for Synthetic B (noting that we did not have Subsets with six to nine tracers).

3.2 NEEM

We now look at the results using the real observations from the NEEM intercomparison by Buizert et al. (2012), with the original five subsets of tracers as used for the synthetic observations. Table 4 shows the weighted RMS mismatch from NEEM observations, Φ_N . All five Subsets used a well-mixed layer for convective mixing near the surface. We also tried

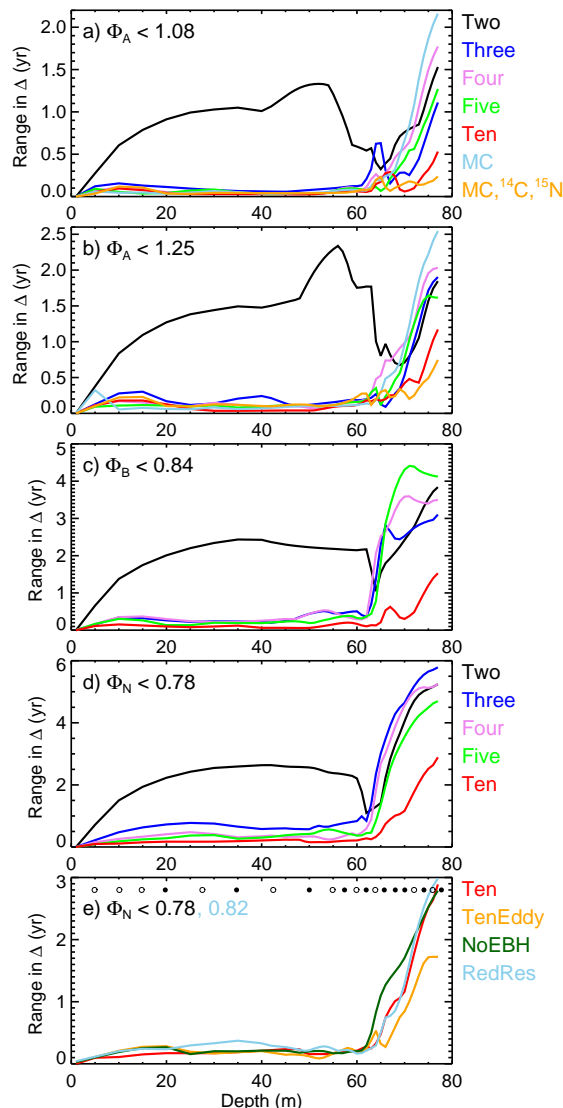


Fig. 5. (a) Depth variation of the range in spectral width of the CO₂ age distribution for Synthetic A representative solutions with $\Phi_A < 1.08$. Cases shown are Subsets Two to Ten, calibration with CH₃CCl₃ alone (denoted “MC” for methyl chloroform), and calibration with CH₃CCl₃, ¹⁴CO₂ and δ^{15} N₂ (denoted “MC, ¹⁴C, ¹⁵N”). Line colours are indicated to the right. (b) Same as in (a) but for $\Phi_A < 1.25$. (c) Range in spectral width versus depth for Synthetic B representative solutions with $\Phi_B < 0.84$ in Subsets Two to Ten, with line colours as in (a). (d) Range in spectral width versus depth for NEEM representative solutions with $\Phi_N < 0.78$ in Subsets Two to Ten, with line colours as in (a). (e) Range in spectral width versus depth for NEEM representative solutions with $\Phi_N < 0.78$ for Subset Ten, case TenEddy and case TenEddy excluding observations corresponding to emissions based histories (denoted “NoEBH”), and with $\Phi_N < 0.82$ for case TenEddy excluding about half of the measurement depths (denoted “RedRes”). Filled circles indicate the measurement depths retained, and open circles indicate the depths discarded, in case RedRes. Line colours are indicated to the right.

Table 4. Weighted RMS mismatch of calculated concentrations for each tracer from the NEEM observations (Φ_N) for the best solution from each of the five Subsets in Table 2 plus Subset Ten with exponential eddy diffusion (denoted TenEddy), Subset Ten with relative diffusion coefficients estimated (TenDC) and TenEddy with relative diffusion coefficients estimated (TenDCEddy).

	Two	Three	Four	Five	Ten	TenEddy
CO ₂	1.37	1.04	0.86	0.90	0.72	0.76
CH ₄	0.46	0.50	0.56	0.67	0.65	0.61
SF ₆	0.69	0.43	0.45	0.51	0.61	0.56
CFC-11	0.94	0.95	0.96	1.01	0.95	0.94
CFC-12	0.92	1.06	1.02	1.15	0.85	0.85
CFC-113	0.70	0.71	0.67	0.76	0.53	0.55
HFC-134a	1.44	1.33	1.01	0.93	0.95	0.95
CH ₃ CCl ₃	1.31	1.11	0.95	0.58	0.74	0.77
¹⁴ CO ₂	0.91	0.83	0.80	1.05	0.72	0.66
$\delta^{15}\text{N}_2$	0.62	0.63	0.64	0.65	0.64	0.60
All	0.98	0.89	0.81	0.84	0.75	0.74
Fitted	0.55	0.53	0.67	0.66	0.75	0.74

Subset Ten with exponentially-decreasing eddy diffusion for convective mixing (denoted TenEddy) instead of the well-mixed layer, and achieved Φ_N that was lower by less than 0.01 than the well-mixed layer case.

We find a significant range (mostly around 0.1–0.2) in the values of Φ_N for individual tracers when we compare solutions with total Φ_N very close to the best solution (within 0.02), in each of the Ten and TenEddy cases. This is due to equifinality, and means that the differences in the fit to individual tracers in Table 4 between cases Ten and TenEddy are not important. If we wished to draw conclusions about the two different ways of modelling convective mixing, it would be important to consider a range of solutions that fit the observations well and not just the single best solution.

The depth of the well-mixed layer was estimated as part of the inversion, and was always well resolved with a clear minimum in Φ_N when $\delta^{15}\text{N}_2$ was used for calibration, but was not resolved at all within the prior range of 2 to 5 m in a calibration that included all tracers except $\delta^{15}\text{N}_2$. We saw this to a lesser extent with the Synthetic A calculations, but the difference is much more dramatic with the real case.

The diffusivity and concentration profiles for TenEddy are shown in Fig. 6, with solutions up to $\Phi_N = 0.78$ (the 68% confidence interval) and an alternative threshold of $\Phi_N = 0.92$ (the range from Buizert et al. (2012) corresponding to a confidence level of almost 100%). With a best value of $\Phi_N = 0.74$ for NEEM TenEddy, our results are now much better than the case with the CSIRO model shown in Buizert et al. (2012) that omitted the upward flux of air due to compression of firn channels and had $\Phi_N = 0.92$. Like the other models in Buizert et al. (2012), we have diffusion (in our case it is molecular) of around $0.1 \text{ m}^2 \text{ yr}^{-1}$ below the lock-in depth of around 63 m. Our modelled $\delta^{15}\text{N}_2$ increases very slightly

through the lock-in zone, at a rate of about 0.0004 \% m^{-1} , which is as consistent with the observations as constant levels. To allow comparison of our results with those from other firn models in Buizert et al. (2012), we include in our Supplement the equivalent model results files as in the Supplement of Buizert et al. (2012).

Figure 7 shows half the range of accepted solutions for NEEM case TenEddy for three different Φ_N thresholds (0.78, 0.8 and 0.92), as well as the observation uncertainties (1σ) and the mismatch between our best solution and the observations. In most cases the model–data mismatch is less than the measurement uncertainty. The spread for the lower threshold (based on the 68% confidence level) is generally about the same as or less than the observation uncertainties. In some cases the spread of solutions is significantly less than the observation uncertainties, such as ¹⁴CO₂ and CH₄ above 60 m and the CFCs above 40 m. The lower threshold for Φ_N is probably the most appropriate choice if we wish to fit to the observations as well as suggested by the uncertainties. However we might choose to use a higher threshold to allow for greater uncertainty due to model error or unknown processes. The spread for the highest threshold is sometimes significantly higher than the observation uncertainties.

Figure 5d shows the variation with depth in the spectral width range from the representative solutions, for the five Subsets of NEEM observations. Compared to Subset Two, Subset Three constrains the spectral width much better above the lock-in depth, but worse below it (using the same threshold for both Subsets, Subset Three can fit the SF₆ observations well within the uncertainties through the diffusive part of the firn leaving scope for larger misfit in the lock-in zone). The range in spectral widths for the NEEM observations is a bit larger than for the Synthetic B observations due to the larger observation uncertainties.

In Fig. 5e we compare the spread in spectral width for case Ten and TenEddy. The spread in TenEddy is larger than Ten below 65 m. This would be offset to some extent by the smaller spread around 5–20 m. We also show a case (denoted “NoEBH”) that is similar to case TenEddy but it excludes firn observations corresponding to atmospheric histories based on emissions estimates (here we exclude SF₆, CH₃CCl₃ and the CFCs below 68 m). Both cases are for $\Phi_N < 0.78$ which corresponds to the 68% confidence level in both cases. We see an increase in the spread below 65 m due to leaving out the observations. We also show a case (denoted “RedRes”, for Reduced Resolution) that uses only 10 observations of each tracer, to test the effect of the depth resolution of reference tracers. The open circles at the top of Fig. 5e show the observation depths that were not used, and the filled circles show the observations depths that were used. We left out observations at roughly every second measurement depth, but discarded an extra depth in the upper firn and retained an extra depth through the region 65–70 m around the peak in ¹⁴CO₂. We show the range for $\Phi_N < 0.82$ which corresponds to the 68% confidence level for the observations

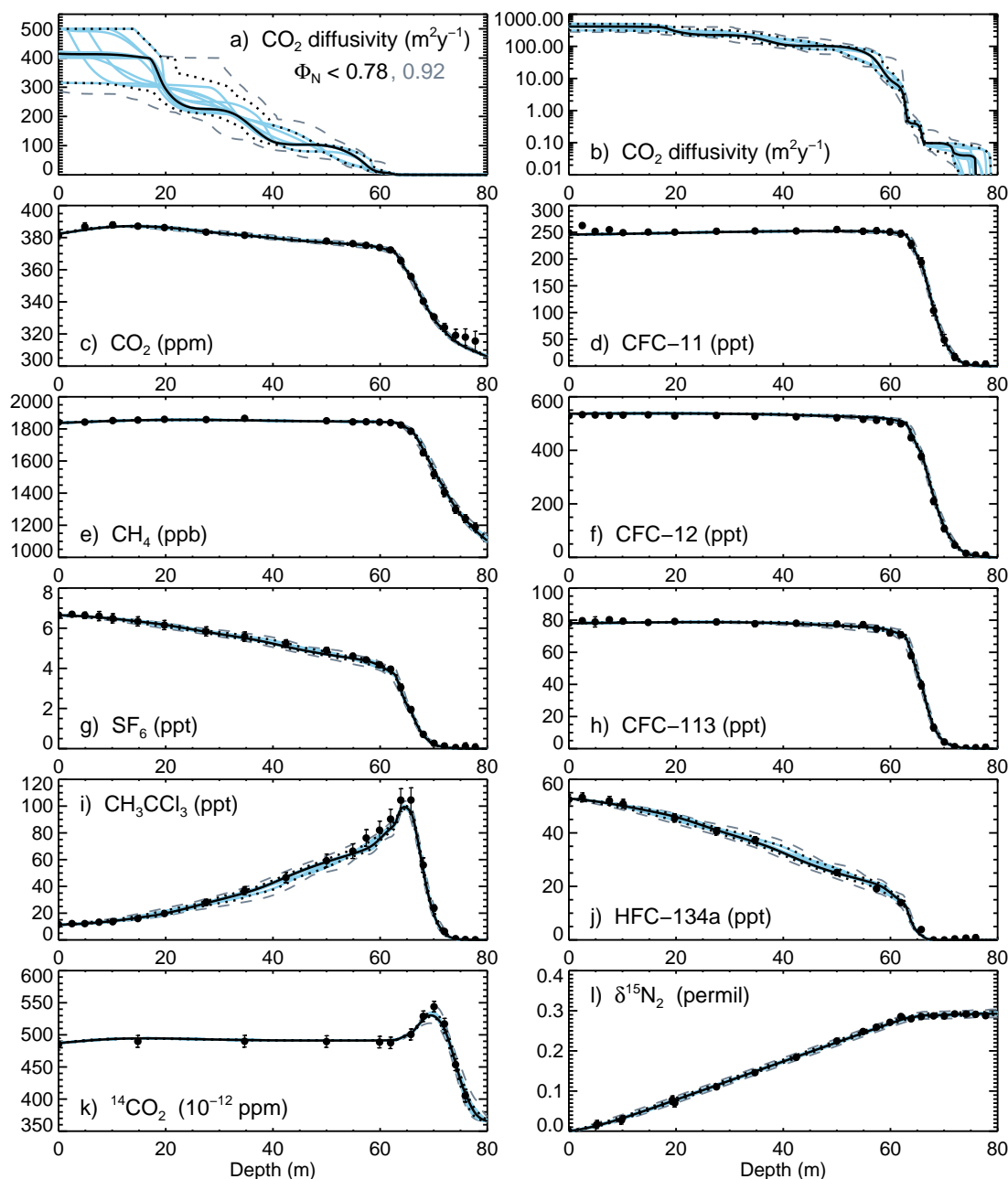


Fig. 6. CO₂ diffusivity and concentration profiles from calibration with all ten NEEM reference tracers from Buizert et al. (2012), with convective mixing modelled by exponentially-decreasing eddy diffusion (TenEddy). Results are for $\Phi_N < 0.78$ except for the blue dashed lines which are for $\Phi_N < 0.92$.

retained (Φ_N increases for the same confidence level due to the smaller number of observations). The spread in spectral width is a bit larger between 25–40 m and below 65 m than case TenEddy, although it seems that most of the increase in spread is due to the higher Φ_N rather than the diffusivity not being as well constrained.

We can get an indication of how useful additional tracers would be for constraining diffusivity at NEEM by running the model in a forward sense with our ensemble of repre-

sentative diffusivity profiles to see how much the modelled concentration profiles vary. We tried this for CCl₄ (which increases in the atmosphere from about 1920 until 1990 followed by a decrease), HCFC-142b (gradually increases from the early 1970s followed by a much more rapid increase from around 1990) and HFC-43-10mee (rapid increase from 2002). Here we are not using firn measurements of these tracers, but instead using the spread in the calculated concentration depth profiles to suggest whether tracers such as these

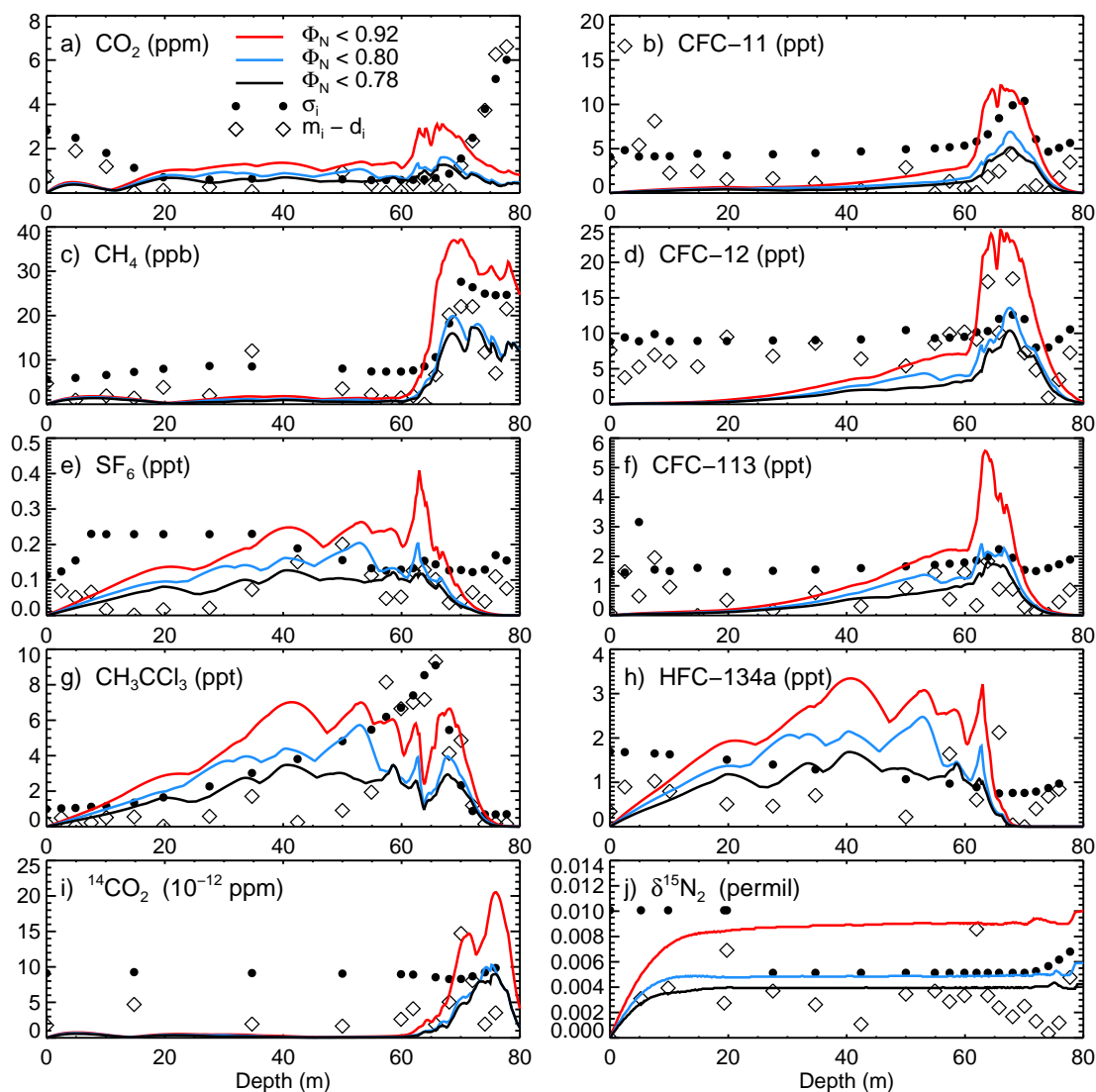


Fig. 7. For each tracer, the solid lines show half the range of all accepted solutions with $\Phi_N < 0.78$ (black), $\Phi_N < 0.8$ (blue) and $\Phi_N < 0.92$ (red) for the case NEEM TenEddy. Filled circles show the observation uncertainties (1σ), and open diamonds show the absolute value of the mismatch between our best solution and the observations.

would be useful for calibration if they were measured accurately enough in the firn and we knew their atmospheric histories well enough. Atmospheric histories, firn depth profiles calculated with diffusivity from NEEM case TenEddy for $\Phi_N < 0.78$ and half the spread for this threshold are shown in Fig. 8. CCl_4 does not have much variation with depth above 60 m, and little spread in the representative solutions, suggesting that it may not add significant new information beyond what we already have from the ten tracers (apart from the degree to which additional observations with independent errors reduce the possibility of fitting to the noise). HCFC-142b has a depth profile similar to HFC-134a, and a similar amount of spread in representative solutions, so this tracer looks like it contains useful, but not unique, information. HFC-43-10mee has a large concentration gradient in the up-

per firn (relative to its overall range), and a significant spread in representative solutions around 20 m. It is unlike the tracers we used for calibration, and probably contains unique information, implying that a tracer with an atmospheric history like this would be worth including for calibration. The value of any tracer depends on how accurately it can be measured and how certain its atmospheric history is, and these uncertainties would need to be less than the range shown in the third column for the tracers to be valuable as additional reference tracers. This type of test is also relevant for predicting how accurately an unknown tracer would be reconstructed, assuming we know approximately how it has changed in the atmosphere (perhaps from emissions estimates). From these calculations, we see that the uncertainty in the firn model (as

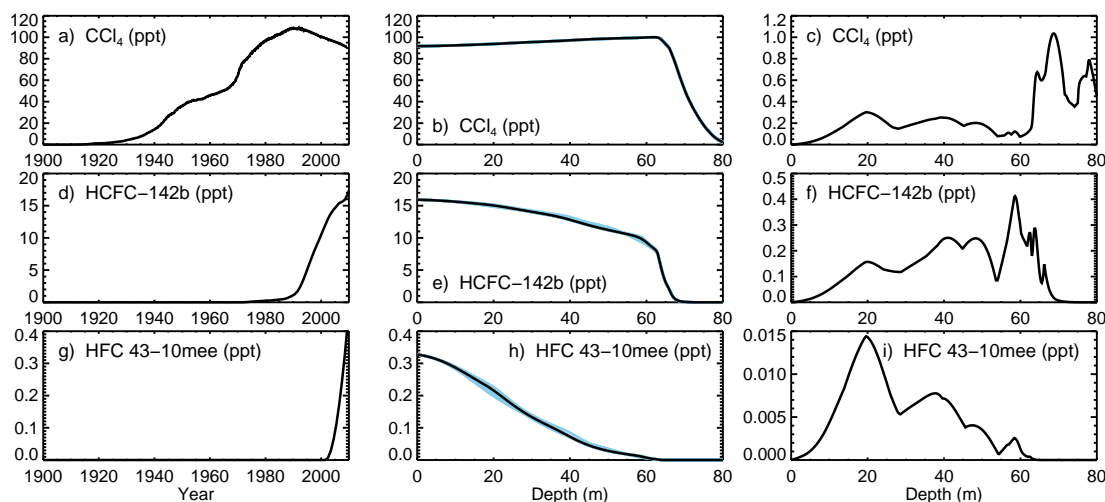


Fig. 8. The left column shows Northern Hemisphere atmospheric histories of CCl_4 , HCFC-142b and HFC-43-10mee (CCl_4 from Martinerie et al. (2009) and HCFC-142b and HFC-43-10mee are Representative Concentration Pathways (Meinshausen et al., 2011) with the RCP3-PD (2.6) from the IMAGE model after 2005). The middle column shows modelled firn profiles at NEEM for these tracers, where the black lines show the best case and the blue lines show 19 representative solutions for $\Phi_N < 0.78$. The right column shows half the spread of representative solutions for this threshold.

it has been calibrated for NEEM) is likely to be less for CCl_4 than for the other two tracers.

3.3 DE08-2

Figure 9 shows results for the DE08-2 firn. The lowest value of Φ_D that the GA finds for these observations is 0.64, and we consider solutions up to $\Phi_D = 0.8$ corresponding to a confidence level of 68 %. The solution with the lowest Φ_D has diffusion reduced by 89 % across the melt layer. Most of the representative solutions have reduction of mixing in the range 85–92 %, however there are two solutions with reduction around 50 %.

The DE08-2 lock-in depth is around 73 m based on the change in slope of the concentration measurements, and our molecular diffusion extends well below this depth. We have some solutions with $\Phi_D < 0.8$ that have molecular diffusion of around $1.0 \text{ m}^2 \text{ yr}^{-1}$ for a significant part of the lock-in zone.

There are strong similarities in the structure of our best case diffusivity for DE08-2 and the diffusivity estimated for DE08-2 by Witrant et al. (2012), particularly the step-like variation with plateaus at around 50 m and 73 m. This is despite the use of different firn models, different functions for diffusivity (including the fact that we use a melt layer and they do not) and different calibration observations (in addition to CO_2 and CH_4 used in both studies, we used $^{14}\text{CO}_2$ and $\delta^{15}\text{N}_2$ and they used CFC-11 and CFC-12).

3.4 DSSW20K

For DSSW20K, we calibrated the firn model with a well-mixed layer for convective mixing near the surface and this gave a best value of $\Phi_W = 1.12$. We also calibrated the model with exponentially-decreasing eddy diffusion for convective mixing near the surface, and this case gave a better fit to the observations with $\Phi_W = 0.92$. However the exponentially-decreasing eddy diffusion profiles preferred by the model lead to significant eddy diffusion extending throughout the firn and into the lock-in zone (this was not the case for NEEM, and could be due to either the shorter firn column at DSSW20K or because eddy diffusion in the lock-in zone improves the fit to observations). As it had not been our intention for eddy diffusion to extend into the lock-in zone, we also tried a third case where we recalibrated using the same form for exponentially-decreasing eddy diffusion above 20 m but with linear extrapolation of eddy diffusion to zero between 20 and 30 m. This achieved $\Phi_W = 0.99$, not as good as the original eddy case, but still better than the well-mixed layer case. This is an interesting result, in terms of recent discussion about the possibility of dispersion in the lock-in zone (see Supplement for further detail). The two cases without eddy diffusion in the lock-in zone had molecular diffusion extending into the lock-in zone. The most noticeable difference between the cases is that with eddy diffusion extending into the lock-in zone there is a better fit to the peak in $^{14}\text{CO}_2$ than for the other cases, although none of them manage to fit the measured $^{14}\text{CO}_2$ at 47 m. We use the third case (with eddy diffusion forced to zero by 30 m) as our preferred case for DSSW20K, and this is shown in Fig. 10. We consider

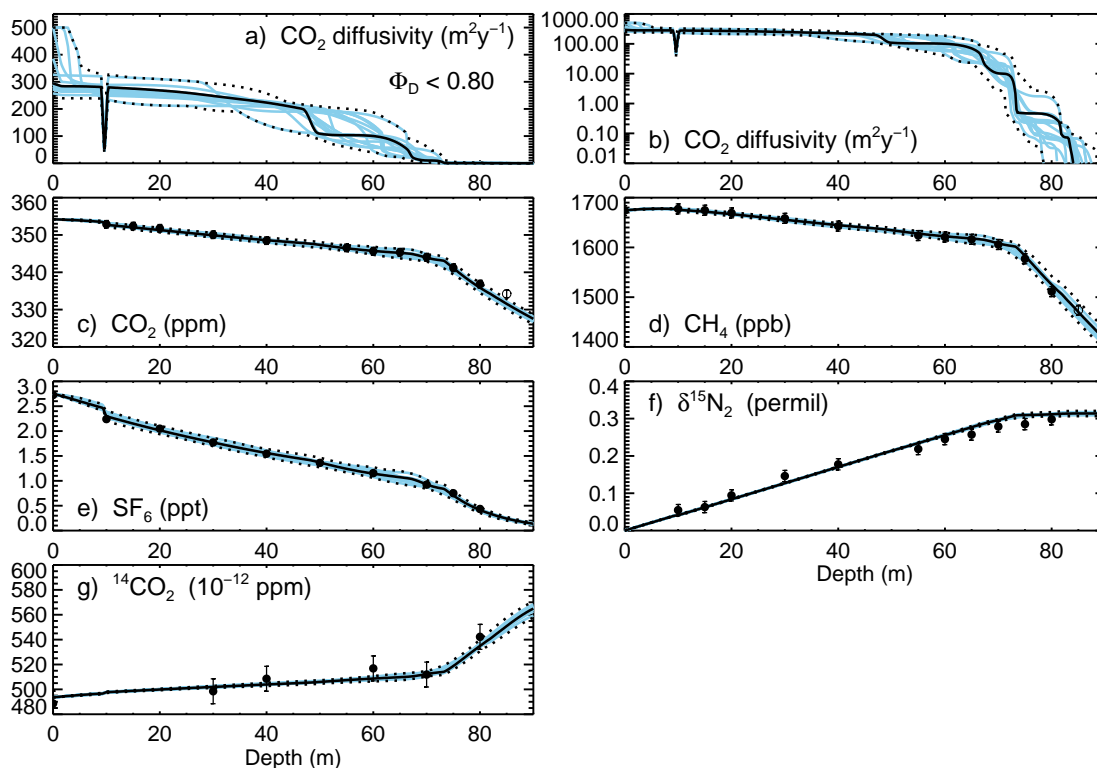


Fig. 9. Diffusivity and tracers at DE08-2 for $\Phi_D < 0.8$.

solutions with Φ_W up to 1.17 corresponding to a confidence level of 68 %.

In Trudinger et al. (2002) and also here, modelled CO_2 at the bottom of the firn is lower than observed, and our best case does not match the $^{14}\text{CO}_2$ peak particularly well. In Sturrock et al. (2002) we had a discrepancy between the DSSW20K firn measurements of CFC-113, CFC-115 and halons H-1211 and H-1301 dated with the old CSIRO firn model and the atmospheric record from the Cape Grim Air Archive. We find that this discrepancy has disappeared with the new version of the CSIRO firn model and our best set of diffusivity parameters for DSSW20K, most likely because we are now including the upward flow of air due to compression. The values of γ_X that we use here differ from the values used in Sturrock et al. (2002), but the difference is too small to explain the discrepancy.

Between about 30–38 m, there is greater spread in the calculated concentrations of many of the tracers than in other parts of the firn. This is most likely a consequence of the gap of more than 10 m between sampling depths for most tracers. Although this region is above the lock-in depth of 43 m, the diffusivity decreases by at least a factor of 10 over this depth range. The sampling depths at DSSW20K were chosen to give priority to measurements in the lock-in zone, to provide most samples of the oldest air while still giving some samples through the diffusive part of the firn to constrain diffusion there. More sampling depths, particularly in the 10 m

above the lock-in zone, would most likely reduce the uncertainty in diffusion in that region.

In the Supplement we look in detail at dispersion in the lock-in zone, and calculate the spread in isotopic fractionation of $\delta^{13}\text{CO}_2$ for the three DSSW20K cases discussed here plus two cases with dispersion in the lock-in zone.

3.5 South Pole

Figure 11 shows results from South Pole 1995. The best Φ_{S95} obtained was 0.92, and we consider solutions up to $\Phi_{S95} = 1.0$ which corresponds to a confidence level of 68 %. We currently do not include thermal diffusion (Severinghaus et al., 2001) in the CSIRO firn model, so have excluded $\delta^{15}\text{N}_2$ observations near the surface that are affected by thermal diffusion at this site (the other sites we model have much less effect from thermal diffusion). Even without using the measurements of SF_6 , CFCs, and CH_3CCl_3 from the lower firn for calibration, the spread of representative solutions for these tracers is quite small, and in good agreement with the observations. These observations could be used with increased uncertainties, however any errors are likely to have systematic rather than random variation with depth, and could also be correlated between species.

Figure 12 shows results from South Pole 2001. Our best Φ_{S01} is 1.24. We consider solutions up to $\Phi_{S01} = 1.38$ corresponding to a confidence level of 68 %.

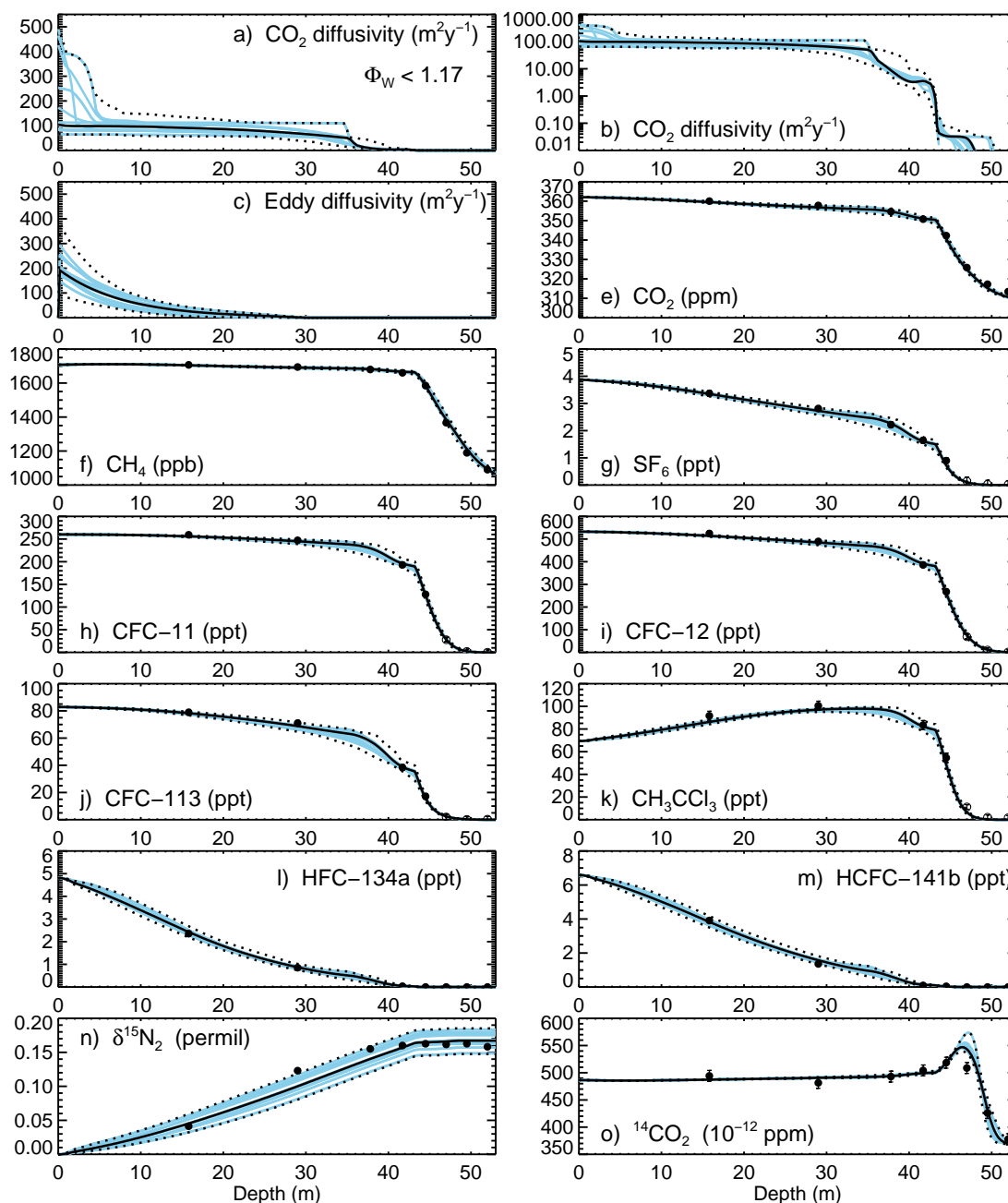


Fig. 10. Diffusivity and tracers at DSSW20K for $\Phi_w < 1.17$. This case has exponentially-decreasing eddy diffusivity for convective mixing near the surface until 30 m then linear decrease to zero. Solid circles show observations used for calibration, open circles show additional measurements that were not used for calibration.

3.6 Comparison of sites

In Fig. 13 we show CO_2 age distributions and spectral width variation with depth for NEEM, DE08-2, DSSW20K, South Pole 1995 and South Pole 2001. The ranges for all sites correspond to a 68 % confidence level (i.e. 1σ). The age distributions for all sites correspond to depths (indicated on the plots) where CO_2 has a mean age of about 1940. The DE08-2 firn only contains air with CO_2 age back to about 1978,

so we have calculated the age distribution in trapped air for DE08-2.

Although we treat South Pole 1995 and 2001 as separate sites, most of the difference between their age distributions in Fig. 13 is probably due to calibrating with different tracers. The age distribution above 50 m is more tightly constrained for South Pole 1995 than South Pole 2001, presumably because of the extra tracers at South Pole 1995. South

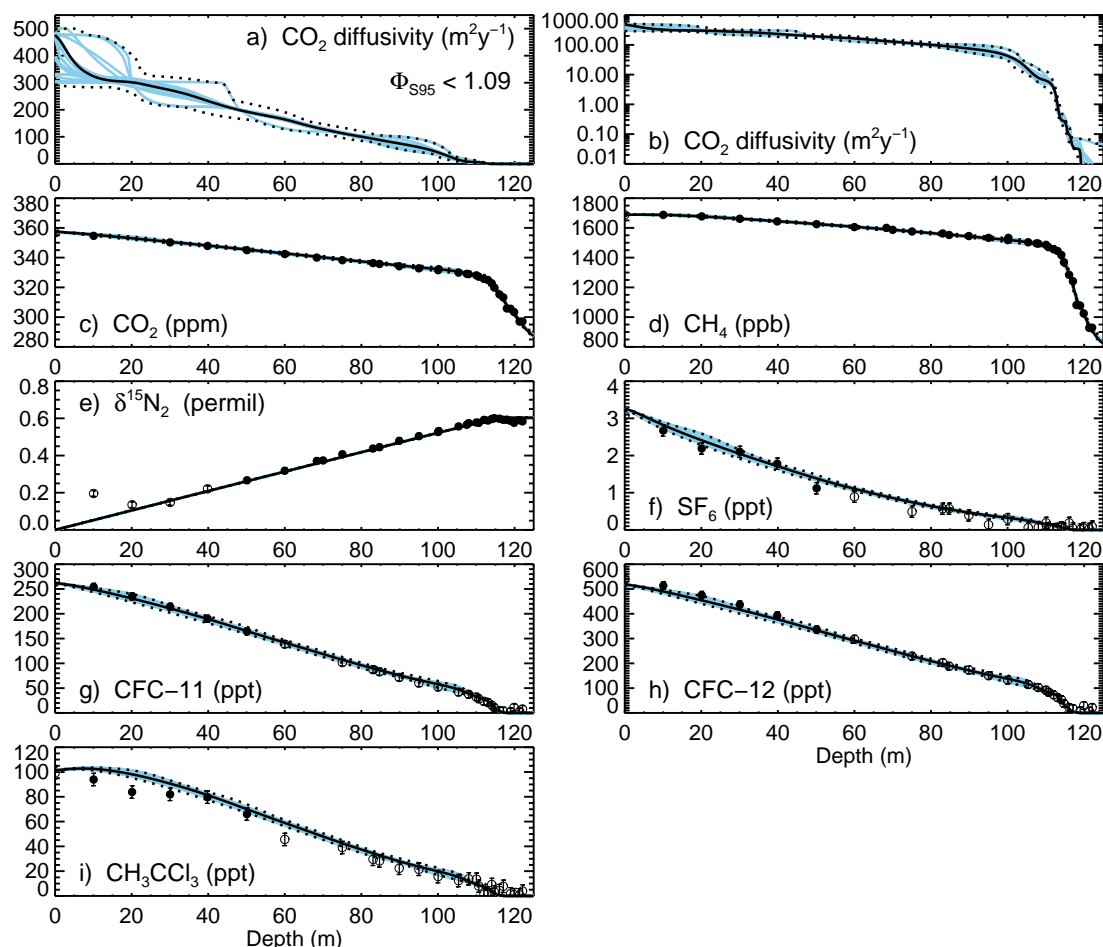


Fig. 11. Diffusivity and tracers at South Pole 1995 for $\Phi_{S95} < 1.09$. Solid circles show observations used for calibration, open circles show additional measurements that were not used for calibration.

Pole 1995 has a wider range of uncertainty below 115 m than South Pole 2001, but the two sites have used similar measurements in this region, so the reason for the difference is not clear.

DSSW20K has a large range in the spectral width at 52 m. The representative solution with the narrowest age spread is the solution with the highest peak in $^{14}\text{CO}_2$ in Fig. 10. This solution fits most observations well, but is significantly higher than the measured $^{14}\text{CO}_2$ at 47 m, which none of the solutions we have found fit well. The DSSW20K case we are showing here has only molecular diffusion in the lock-in zone, and the spread of spectral width may increase if we included solutions with eddy diffusion in the lock-in zone.

Of the sites we have modelled, DE08-2 ice has the narrowest age spread at 1940 (with a best estimate of the spectral width of 4.5 yr and a 1σ range of 4.4 to 4.7 yr), followed by DSSW20K firn (6.0 yr with a 1σ range of 3.6–6.9 yr), NEEM firn (10.8 yr with a 1σ range of 9.6–11.3 yr) then South Pole firn (the 1995 site has a best estimate of 16.0 yr with a 1σ range of 15.7–18.9 yr and the 2001 site has a best estimate

of 18.1 yr and a 1σ range of 17.2–18.2 yr). The uncertainty range shown for DE08-2 is due only to our representative diffusivity profiles, and does not include uncertainty in bubble trapping due to uncertainty in the closed porosity profile or other factors influencing trapping. Tests with a range of alternative open porosity profiles to include this contribution to the uncertainty in trapping, where we retune diffusivity as a function of open porosity to match observations in each case, showed a greater range for the spectral width, as expected. However, the total range is still below all estimates for the other sites apart from DSSW20K. Due to the high accumulation rate at DE08-2, bubble trapping doesn't increase the spectral width for trapped air significantly beyond what already occurs in the firn, leaving DE08-2 with probably the narrowest spectral width of the sites we consider.

Modelled CO_2 at NEEM, DSSW20K and DE08-2 deep in the firn is lower than measured. It has been known for some time that northern hemisphere CO_2 may be influenced by in situ production (Anklin et al., 1995), however this was not expected for Antarctic sites. This needs further work to

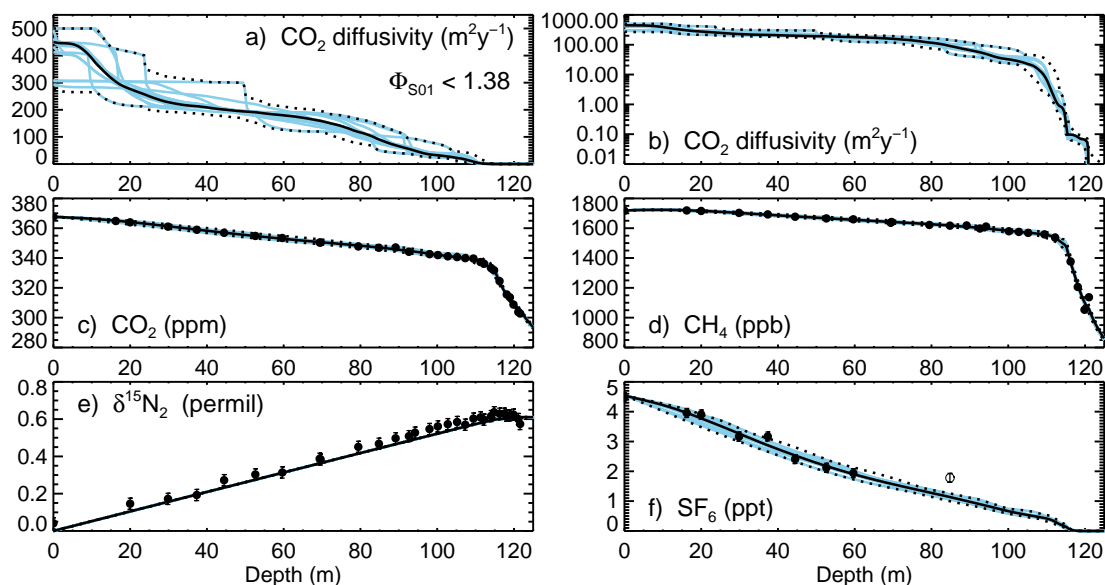


Fig. 12. Diffusivity and tracers at South Pole 2001 for $\Phi_{S01} < 1.38$. Solid circles show observations used for calibration, open circles show additional measurements that were not used for calibration.

understand, and the isotopes of CO_2 may help. We have not included any contribution to $^{14}\text{CO}_2$ from ^{14}C produced in situ in the ice matrix by cosmic rays being released from the ice matrix to firn air (Smith et al., 2000). This process is currently very uncertain. The reasonable agreement our modelled $^{14}\text{CO}_2$ profiles show with the measurements suggest that in situ production does not have a large influence. However, it could still contribute to the observed variation, so firn model calibration using $^{14}\text{CO}_2$ might be improved by better understanding of this process.

In the Supplement, we calibrated the relative diffusion coefficients γ_X in addition to the other diffusivity parameters, for both synthetic datasets and NEEM, DE08-2 and DSSW20K. In the synthetic calculations we do reasonably well with Synthetic A, but have serious concerns with Synthetic B which leads us to be cautious of the results for the real observations. While uncertainties exist in γ_X , the best approach is probably to use the best independent estimates for γ_X , but to include some representative solutions with different values of γ_X to account for this uncertainty.

We have applied our model and calibration technique to several sites with quite different properties and tracers available for calibration. All sites that we modelled required some diffusion in the lock-in zone.

4 Discussion

We have seen that some tracers constrain diffusivity more than others. In particular, tracers with more variation through the diffusive part of the firn, like CH_3CCl_3 , HFC-134a, SF_6 and HFC-43-10mee at NEEM, are most useful for constrain-

ing diffusivity in this region. Similarly, tracers with most variation in the lock-in zone, like $^{14}\text{CO}_2$ (at NEEM), CO_2 and CH_4 , constrain diffusion most there. Tracers with constant concentration over parts of the firn are not so useful in those regions (such as CH_4 through the diffusive part). These conclusions do not account for data errors, and the larger the error is for a particular tracer, the less useful it becomes, particularly if the noise is correlated.

Using a larger number of tracers does constrain diffusivity more tightly. In terms of the signal, it appears that three carefully chosen tracers will do nearly as well as ten tracers in constraining (only) molecular diffusivity if the noise is small and Gaussian, and if we assume a monotonic diffusivity. Small improvements may occur with each different tracer that is added. We can quantify how much additional constraint, and for which regions of the firn, is provided by each extra tracer by comparing the spread of solutions with and without the extra tracers. However, when we allow for eddy diffusion, more tracers covering a range of γ_X are needed to distinguish between molecular and eddy diffusion.

In terms of the noise, obviously the larger the noise the more it will overwhelm the signal, and systematic errors are more difficult to deal with than random errors. In the synthetic calculations, we chose data errors to be proportional to the range of concentrations covered by the firn measurements. Actual data errors will vary widely relative to this range, and this will affect which tracers have more information on firn processes. Overfitting noisy data was a problem for up to three tracers with a small amount of normally-distributed noise, and more of a problem with systematic noise. If systematic noise is related for different tracers, overfitting the data will be harder to avoid.

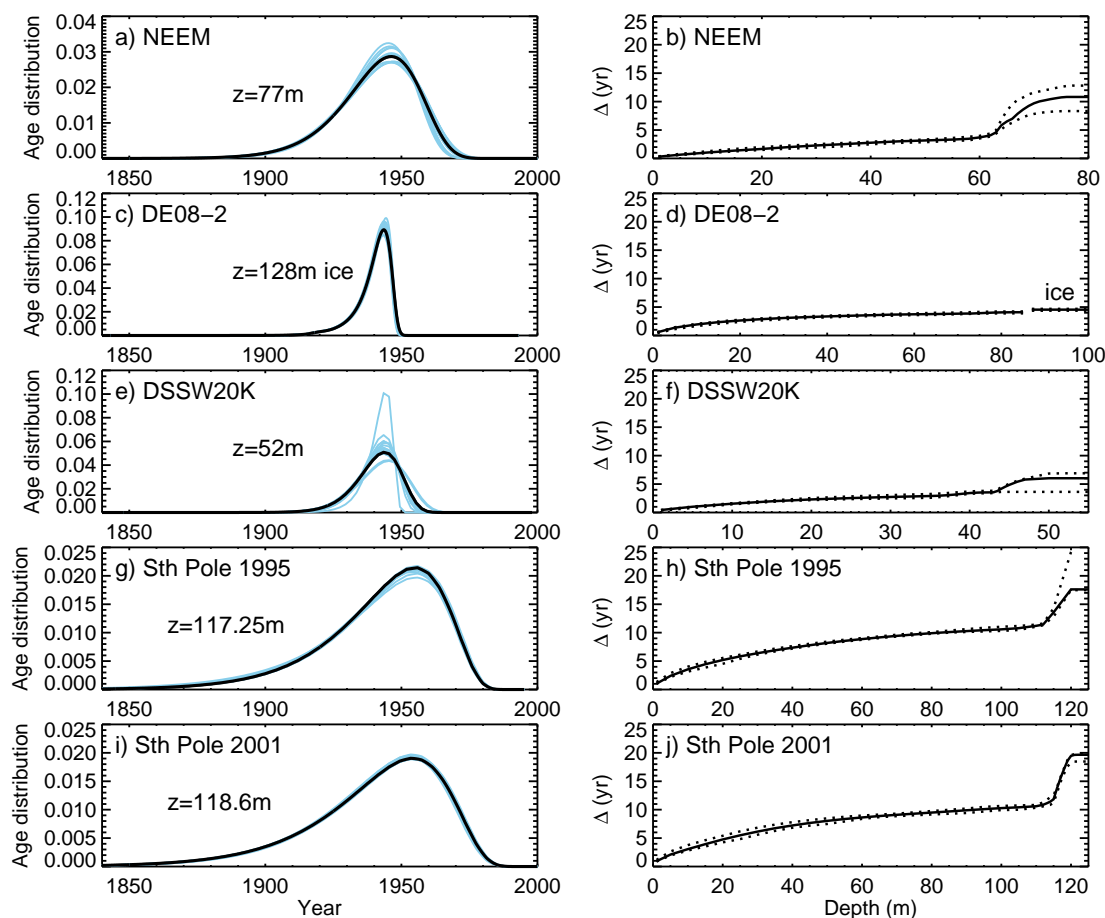


Fig. 13. Plots on the left show CO_2 age distributions for NEEM, DE08-2, DSSW20K, South Pole 1995 and South Pole 2001 corresponding to CO_2 mean ages of 1940. The black lines show the best solution in each case (corresponding to the case with closest match to the reference tracers), and the blue lines show results from 19 representative solutions with diffusivity and other model parameters within the 68 % confidence interval for each site. Plots on the right show variation with depth of the spectral width of the CO_2 age distributions, Δ . The black solid lines are the best case and the dotted lines show the upper and lower ranges of estimates from the 19 representative solutions corresponding to confidence intervals of 68 %. The age distribution for DE08-2 is for air trapped in bubbles in ice at 128 m, and the spectral width plot shows values for the firn to 85 m then ice below this. All other cases are for firn air only.

What matters for a particular site is the timing of features in the atmospheric records relative to the range of air ages in the firn column at the sampling date, so the relative value of different tracers is expected to vary for different sites and drilling dates. For example, the bomb pulse in $^{14}\text{CO}_2$ and peak in CH_3CCl_3 can be very useful signals, but that depends on where they are in the firn, such as in firn air above or below the lock-in depth, or in ice bubbles. At some time in the future, the peak in CH_3CCl_3 will move into the lock-in zone for some sites, so will then provide a useful tracer to constrain lock-in zone diffusion. It will eventually move into trapped bubbles in ice, where it might be useful to help constrain trapping in addition to firn diffusion. Tracers that have been useful in the past for constraining firn diffusivity may not be as useful in the future, but the principles for choosing tracers will remain the same.

The fact that the range of solutions is greater for tracers with unique information implies that it may be worth looking more closely at the weights of different observations, so that independent pieces of information are considered more equally, rather than simply using the observation uncertainties. This is an issue because the cost function treats all observations as independent, but model errors for tracers with similar histories would be correlated (e.g. as the CFCs have similar histories, the impact of higher diffusivity in a region of the firn will be similar). We might wish to downweight tracers with similar histories (e.g. CFCs) relative to tracers with unique information (e.g. $^{14}\text{CO}_2$), to reflect this model error correlation. In our NEEM calculations, we had many tracers with significant information through the diffusive region, but only 3 tracers with significant information through the lock-in zone (CO_2 , CH_4 and $^{14}\text{CO}_2$), and we saw greater spread in these tracers in cases where all ten tracers were used. A more

equal weighting between different types of tracers might improve the calculation.

Based on our study, the recommendation for future firn campaigns would be to give priority to tracers that have most variation with depth (relative to their uncertainties) in different parts of the firn. $\delta^{15}\text{N}_2$ is useful for constraining parameters related to convective mixing, and CH_3CCl_3 and $^{14}\text{CO}_2$ have been particularly useful at NEEM for the diffusive, and lock-in zones, respectively. The more tracers that can be measured, and with good characterisation of their uncertainties, the less chance there will be of overfitting noise. There is value in making as many measurements as possible through the firn column, with particular emphasis on the lock-in zone and the 10 m or so above the lock-in depth where diffusivity changes rapidly with depth.

We modelled convective mixing near the surface at NEEM and DSSW20K in two different ways – with a well-mixed layer and exponentially-decreasing eddy diffusion. For NEEM, the overall fit to calibration observations was very similar (difference in Φ of less than 0.01), whereas for DSSW20K the difference was greater, with the exponential eddy diffusion form always slightly better. Cases using each of these forms could be included in a representative set of diffusivity parameters.

Most of our calculations involved estimating only molecular diffusivity and a depth for the well-mixed layer. In this case we found that the uncertainty due to data uncertainties was more important than equifinality beyond a few carefully selected tracers (as long as we allow only monotonic diffusivity profiles). However, when we added additional processes such as dispersion in an attempt to take account of uncertainties in model structure, equifinality became more important. In that case, different types of tracers might be needed to reduce the equifinality (e.g. isotopic ratios might be the best way to learn about dispersion in the lock-in zone).

How important the uncertainties are depends on what we are doing with the model. Reconstructing the atmospheric history of a tracer that has a similar history to a tracer used for calibration would be expected to have a small uncertainty. In contrast, an estimate of the diffusive fractionation affecting isotopic ratios might have large uncertainty (we saw this for $\delta^{13}\text{CO}_2$ at DSSW20K). As the number of observations increases, the confidence level associated with a chosen threshold increases, but if important processes are missing from the model, we may seriously under-estimate the errors in a model prediction.

The firn model intercomparison by Buizert et al. (2012) showed one result each from six different firn models, with some variation in the processes that were included, most notably dispersion in the lock-in zone. Here we looked at multiple solutions compatible with the observations over the same range of model-data mismatch with only one firn model. The range of results we obtained for NEEM case TenEddy for a similar range of Φ was not far from that obtained with the six models. The range in spectral width of the CO_2 age dis-

tribution at 78 m in Buizert et al. (2012) was 8.3–12.5 yr, compared to the range we obtain at the same depth (with $\Phi_{\text{N}} < 0.92$) of 8.4–12.8 yr (without dispersion). Buizert et al. (2012) calculated a range in their Diagnostic 1 (the diagnostic that quantifies isotopic fractionation) for the six models of about 0.04 ‰ at 70 m and 0.05 ‰ at 80 m. Our NEEM TenEddy range for this diagnostic was about 0.025 ‰ at 70 m and 0.04 ‰ at 80 m. Our range is smaller than the range in Buizert et al. (2012), but still significant considering our cases had no dispersion in the lock-in zone and Buizert et al. (2012) attributed the range in Diagnostic 1 to the different ways lock-in zone diffusion was parameterised in the models.

The uncertainty ranges from our study and Buizert et al. (2012) are both important, but not independent. If we include the equifinality due to uncertain processes like dispersion in one firn model it should cover some of the range covered by different models, but there may still be other differences between models that will not be captured with one model alone. Part of the difference between the models in Buizert et al. (2012) will be due to the fact that a single solution from each model was compared, whereas we would expect that there will be multiple solutions that fit the observations almost as well as the best cases, as we have found here, and these will probably have overlapping ranges to some extent. However, both this work and the Buizert et al. (2012) model intercomparison suggest that the use of a single calculation from one firn model for reconstruction of atmospheric records is not adequate to capture the uncertainty in model outputs caused by parameter equifinality and uncertainty in firn processes.

5 Summary and conclusions

We have developed a new version of the CSIRO firn model and used it to test how well different tracers constrain the firn diffusivity profile, one of the most important and uncertain of the required model inputs. Our method for reconstructing the profile of molecular diffusivity in firn based on the genetic algorithm and monotonic splines was successful in two synthetic cases with different error characteristics. When calibrating the model, we generate an ensemble of representative diffusivity parameter sets that all fit the calibration data adequately. This ensemble represents the uncertainty in the model due to data uncertainty, equifinality and in some cases uncertainty in model processes. When we compare solutions that fit the observations almost as well as the best solution, we often see significant variation in the fit to individual tracers due to equifinality, and this should be taken into account when the results from different firn models or for different model configurations are compared.

The best tracers for calibrating molecular diffusivity in firn are those with most variation with depth relative to their uncertainties (where their uncertainty includes the uncertainty in their atmospheric history). In a synthetic case for NEEM

with small, Gaussian errors, CH_3CCl_3 and HFC-134a were particularly useful above the lock-in zone and $^{14}\text{CO}_2$ below the lock-in zone. $\delta^{15}\text{N}$ is useful for constraining parameters related to convective mixing near the surface. It appears that three carefully selected tracers will do almost as well as a larger number of tracers in terms of constraining molecular diffusivity alone when data errors are small and Gaussian, however with either realistic data errors or the need to constrain additional processes there is benefit to including as many tracers as possible. Overfitting of noisy data is a problem even for purely random noise and a more serious problem for systematic errors.

We have applied our model and calibration technique to several sites with quite different properties and measured tracers available for calibration. All of the sites we modelled had non-zero diffusion below the lock-in depth, of the order of $0.1\text{--}1.0\text{ m}^2\text{ yr}^{-1}$. The inclusion of the upward flow of air due to compression in the new version of the CSIRO firn model is clearly important. Although diffusivity can compensate for its absence to some extent, we are now able to match observations better with the upward flow.

This work has been an attempt to quantify uncertainties due mainly to data uncertainties and equifinality. Our calculations for convective mixing and dispersion in the lock-in zone are a start to including the effects of uncertainty in model processes, but further work is needed to define the best functions and parameter ranges to use based on our best knowledge of the physics (e.g. moving beyond the exponential function for convective mixing near the surface), and to include other uncertain processes (such as in situ production of CO_2 or ^{14}C). We challenge the idea of using a firn model with a single parameter set, instead suggesting the use of multiple parameter sets, preferably with multiple representations of uncertain processes. Uncertainty due to equifinality does reduce as more tracers are used for model calibration, but it does not disappear completely, particularly if we allow for uncertain processes like dispersion in the lock-in zone.

Supplementary material related to this article is available online at: <http://www.atmos-chem-phys.net/13/1485/2013/acp-13-1485-2013-supplement.zip>.

Acknowledgements. This work has been undertaken as part of the Australian Climate Change Science Program, funded jointly by the Department of Climate Change and Energy Efficiency, the Bureau of Meteorology and CSIRO. CSIRO's research in this area was also funded by the Centre for Ice and Climate at the University of Copenhagen. The Centre of Excellence of Mathematics and Statistics of Complex Systems (MASCOS) has been funded by the Australian Research Council. I. Enting's fellowship at MASCOS was funded in part by CSIRO. P. Rayner is in receipt of an Australian Professorial Fellowship (DP1096309). C. Buizert was

supported by the NOAA Climate and Global Change Fellowship Program, administered by the University Corporation for Atmospheric Research. We would like to thank all those involved in the logistics, drilling, sampling and analysis for the firn air campaigns at DE08-2, South Pole 1995, DSSW20K, South Pole 2001 and NEEM 2008. We thank Patricia Martinerie for useful discussions and sharing atmospheric histories, Jeff Severinghaus for advice on equations for eddy diffusion and Mark Battle for providing data and advice for the South Pole firn sites. We acknowledge Ben Miller (now at CIRES/University of Colorado) for the Cape Grim Air Archive measurements of HFC-141b used in this work. We thank Jeff Severinghaus and an anonymous reviewer for helpful comments and suggestions.

Edited by: R. van de Wal

References

- Anklin, M., Barnola, J.-M., Schwander, J., Stauffer, B., and Raynaud, D.: Processes affecting the CO_2 concentrations measured in Greenland ice, *Tellus B*, 47, 461–470, doi:10.1034/j.1600-0889.47.issue4.6.x, 1995.
- Árnadóttir, T. and Segall, P.: The 1989 Loma Prieta earthquake imaged from inversion of geodetic data, *J. Geophys. Res.*, 99, 21835–21855, 1994.
- Aydin, M., Saltzman, E. S., De Bruyn, W. J., Montzka, S. A., Butler, J. H., and Battle, M.: Atmospheric variability of methyl chloride during the last 300 years from an Antarctic ice core and firn air, *Geophys. Res. Lett.*, 31, L02109, doi:10.1029/2003GL018750, 2004.
- Battle, M., Bender, M., Sowers, T., Tans, P. P., Butler, J. H., Elkins, J. W., Ellis, J. T., Conway, T., Zhang, N., Lang, P., and Clarke, A. D.: Atmospheric gas concentrations over the past century measured in air from firn at the South Pole, *Nature*, 383, 231–235, 1996.
- Beven, K.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18–36, 2006.
- Buizert, C., Martinerie, P., Petrenko, V. V., Severinghaus, J. P., Trudinger, C. M., Witrant, E., Rosen, J. L., Orsi, A. J., Rubino, M., Etheridge, D. M., Steele, L. P., Hogan, C., Laube, J. C., Sturges, W. T., Levchenko, V. A., Smith, A. M., Levin, I., Conway, T. J., Dlugokencky, E. J., Lang, P. M., Kawamura, K., Jenk, T. M., White, J. W. C., Sowers, T., Schwander, J., and Blunier, T.: Gas transport in firn: multiple-tracer characterisation and model intercomparison for NEEM, Northern Greenland, *Atmos. Chem. Phys.*, 12, 4259–4277, doi:10.5194/acp-12-4259-2012, 2012.
- Butler, J. H., Battle, M., Bender, M. L., Montzka, S. A., Clarke, A. D., Saltzman, E. S., Sucher, C. M., Severinghaus, J. P., and Elkins, J. W.: A record of atmospheric halocarbons during the twentieth century from polar firn, *Nature*, 399, 749–755, 1999.
- Butler, J. H., Montzka, S. A., Battle, M., Clarke, A. D., Mondeel, D. J., Lind, J. A., Hall, B. D., and Elkins, J. W.: Collection and Analysis of Firn Air from the South Pole, 2001, AGU Fall Meeting Abstracts, p. F145, 2001.
- Chen, N. H. and Othmer, D. F.: New generalized equation for gas diffusion coefficient, *J. Chem. Eng. Data*, 7, 37–41, 1962.

- Colbeck, S. C.: Air movement in snow due to windpumping, *J. Glaciol.*, 35, 209–213, 1989.
- Craig, H., Horibe, Y., and Sowers, T.: Gravitational separation of gases and isotopes in polar ice caps, *Science*, 242, 1675–1678, 1988.
- Draper, N. R. and Smith, H.: *Applied Regression Analysis*, Wiley, New York, USA, 709 pp., 1981.
- Etheridge, D. M. and Wooley, C. W.: Ice core drilling at a high accumulation area of Law Dome, Antarctica, 1987, in: *Ice Core Drilling, Proceedings of the Third International Workshop on Ice Core Drilling Technology*, Grenoble, France, October 10–14 1988, edited by: Rado, C. and Beaudoin, D., 86–96, Centre Natl. de la Rech. Sci., Grenoble, France, 1989.
- Etheridge, D. M., Steele, L. P., Langenfelds, R. L., Francey, R. J., Barnola, J. M., and Morgan, V. I.: Natural and anthropogenic changes in atmospheric CO₂ over the last 1000 years from air in Antarctic ice and firn, *J. Geophys. Res.*, 101, 4115–4128, 1996.
- Etheridge, D. M., Steele, L. P., Francey, R. J., and Langenfelds, R. L.: Atmospheric methane between 1000 AD and present: evidence for anthropogenic emissions and climate variability, *J. Geophys. Res.*, 103, 15979–15993, 1998.
- Fabre, A., Barnola, J. M., Arnaud, L., and Chappellaz, J.: Determination of gas diffusivity in polar firn: comparison between experimental measurements and inverse modeling, *Geophys. Res. Lett.*, 27, 557–560, 2000.
- Freitag, J., Dobrindt, U., and Kipfstuhl, J.: A new method or predicting transport properties of polar firn with respect to gases in the pore-space scale, *Ann. Glaciol.*, 35, 538–544, 2002.
- Freitag, J., Wilhelms, F., and Kipfstuhl, S.: Microstructure-dependent densification of polar firn derived from X-ray microtomography, *Ann. Glaciol.*, 50, 243–250, 2004.
- Fuller, E. N., Schettler, P. D., and Giddings, J. C.: A new method for prediction of binary gas-phase diffusion coefficients, *Ind. Eng. Chem.*, 58, 19–27, 1966.
- Hall, T. M., and Plumb, R. A.: Age as a diagnostic of stratospheric transport, *J. Geophys. Res.*, 99, 1059–1070, 1994.
- Haupt, R. and Haupt, S.: *Practical Genetic Algorithms*, John Wiley and Sons, New York, USA, 177 pp., 1998.
- Højberg, A. L. and Refsgaard, J. C.: Model uncertainty-parameter uncertainty versus conceptual models, *Water Sci. Technol.*, 52, 177–186, 2005.
- Hörhold, M. W., Laepple, T., Freitag, J., Bigler, M., Fischer, H., and Kipfstuhl, S.: On the impact of impurities on the densification of polar firn, *Earth Planet. Sci. Lett.*, 325, 93–99, 2012.
- 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, 61–73, 2006.
- Kawamura, K., Severinghaus, J. P., Ishidoya, S., Sugawara, S., Hashida, G., Motoyama, H., Fujii, Y., Aoki, S., and Nakazawa, T.: Convective mixing of air in firn at four polar sites, *Earth Planet. Sci. Lett.*, 244, 672–682, doi:10.1016/j.epsl.2006.02.017, 2006.
- Levchenko, V. A., Etheridge, D. M., Francey, R. J., Trudinger, C. M., Tuniz, C., Lawson, E. M., Smith, A. M., Jacobsen, G. E., Hua, Q., Hotchkis, M. A. C., Fink, D., Morgan, V., and Head, J.: Measurements of the ¹⁴CO₂ bomb pulse in firn and ice at Law Dome, Antarctica, *Nucl. Instr. Meth. B*, 123, 290–295, 1997.
- Lugg, G. A.: Diffusion coefficients of some organic and other vapor in air, *Anal. Chem.*, 40, 1072–1077, 1968.
- Martinerie, P., Nourtier-Mazauric, E., Barnola, J.-M., Sturges, W. T., Worton, D. R., Atlas, E., Gohar, L. K., Shine, K. P., and Brasseur, G. P.: Long-lived halocarbon trends and budgets from atmospheric chemistry modelling constrained with measurements in polar firn, *Atmos. Chem. Phys.*, 9, 3911–3934, doi:10.5194/acp-9-3911-2009, 2009.
- Matsunaga, N., Hori, M., and Nagashima, A.: Mutual diffusion coefficients of halogenated hydrocarbon refrigerant-air systems, *High Temp.-High Press*, 25, 185–192, 1993.
- Matsunaga, N., Hori, M., and Nagashima, A.: Diffusion coefficients of global warming gases into air and its component gases, *High Temp.-High Press*, 30, 77–83, 1998.
- Matsunaga, N., Hori, M., and Nagashima, A.: Measurements of the mutual diffusion coefficients of gases by the Taylor method, 7th report, measurements on the SF₆-air, SF₆-N₂, SF₆-O₂, CFC12-N₂, CFC12-O₂, HCFC22-N₂ and HCFC22-O₂ systems, *Trans. Jpn. Soc. Mech. Eng. B.*, 68, 550–555, 2002.
- Meinshausen, M., Smith, S., Calvin, K., Daniel, J., Kainuma, M., Lamarque, J.-F., Matsumoto, K., Montzka, S., Raper, S., Riahi, K., Thomson, A., Velders, G., and van Vuuren, D.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, *Climatic Change*, 109, 213–241, doi:10.1007/s10584-011-0156-z, 2011.
- Rommelaere, V., Arnaud, L., and Barnola, J.: Reconstructing recent atmospheric trace gas concentrations from polar firn and bubble ice data by inverse methods, *J. Geophys. Res.*, 102, 30069–30083, 1997.
- Schwander, J.: The transformation of snow to ice and the occlusion of gases, in: *The Environmental Record in Glaciers and Ice Sheets*, edited by: Oeschger, H. and Langway, C. C., John Wiley, New York, USA, 53–67, 1989.
- Schwander, J., Stauffer, B., and Sigg, A.: Air mixing in firn and the age of the air at pore close-off, *Ann. Glaciol.*, 10, 141–145, 1988.
- Schwander, J., Barnola, J. M., Andrié, C., Leuenberger, M., Ludin, A., Raynaud, D., and Stauffer, B.: The age of the air in the firn and the ice at Summit, Greenland, *J. Geophys. Res.*, 98, 2831–2838, 1993.
- Seber, G. A. F. and Wild, C. J.: *Nonlinear Regression*, Wiley, doi:10.1002/0471725315, 2005.
- Severinghaus, J. P., Grachev, A., and Battle, M.: Thermal fractionation of air in polar firn by seasonal temperature gradients, *Geochem. Geophys. Geosyst.*, 2, 1048, doi:10.1029/2000GC000146, 2001.
- Severinghaus, J. P., Albert, M. R., Courville, Z. R., Fahnestock, M. A., Kawamura, K., Montzka, S. A., Mühle, J., Scambos, T. A., Shields, E., Shuman, C. A., Suwa, M., Tans, P., and Weiss, R. F.: Deep air convection in the firn at a zero-accumulation site, Central Antarctica, *Earth Planet. Sci. Lett.*, 293, 359–367, doi:10.1016/j.epsl.2010.03.003, 2010.
- Smith, A. M., Levchenko, V. A., Etheridge, D. M., Lowe, D. C., Hua, Q., Trudinger, C. M., Zoppi, U., and Elcheikh, A.: In search of in-situ radiocarbon in Law Dome ice and firn, *Nucl. Instr. Meth. B*, 172, 610–622, 2000.
- Sowers, T., Bernard, S., Aballain, O., Chappellaz, J., Barnola, J., and Marik, T.: Records of the δ¹³C of atmospheric CH₄ over the last 2 centuries as recorded in Antarctic snow and ice, *Global*

- Biogeochem. Cy., 19, GB2002, doi:10.1029/2004GB002408, 2005.
- Sturrock, G. A., Etheridge, D. M., Trudinger, C. M., Fraser, P. J., and Smith, A. M.: Atmospheric histories of halocarbons from analysis of Antarctic firn air: Major Montreal Protocol species, *J. Geophys. Res.*, 107, 4765, doi:10.1029/2002JD002548, 2002.
- Trudinger, C. M.: The carbon cycle over the last 1000 years inferred from inversion of ice core data, PhD thesis, Monash University, available at: http://www.cmar.csiro.au/e-print/open/trudinger_2001a0.htm, 2000.
- Trudinger, C. M., Enting, I. G., Etheridge, D. M., Francey, R. J., Levchenko, V. A., Steele, L. P., Raynaud, D., and Arnaud, L.: Modeling air movement and bubble trapping in firn, *J. Geophys. Res.*, 102, 6747–6763, 1997.
- Trudinger, C. M., Etheridge, D. M., Rayner, P. J., Enting, I. G., Sturrock, G. A., and Langenfelds, R. L.: Reconstructing atmospheric histories from measurements of air composition in firn, *J. Geophys. Res.*, 107, 4780, doi:10.1029/2002JD002545, 2002.
- van der Kemp, W. J. M., Alderliesten, C., van der Borg, K., Holmlund, P., de Jong, A. F. M., Karlöf, L., Lamers, R. A. N., Oerlemans, J., Thomassen, M., and van de Wal R. S. W. Very little in situ produced radiocarbon retained in accumulating Antarctic ice. *Nucl. Inst. Meth. Phys. Res. B*, 172, 632–636, 2000.
- Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39, 1201, doi:10.1029/2002WR001642, 2003.
- Witrand, E., Martinerie, P., Hogan, C., Laube, J. C., Kawamura, K., Capron, E., Montzka, S. A., Dlugokencky, E. J., Etheridge, D., Blunier, T., and Sturges, W. T.: A new multi-gas constrained model of trace gas non-homogeneous transport in firn: evaluation and behavior at eleven polar sites, *Atmos. Chem. Phys.* 12, 11465–11483, doi:10.5194/acp-12-11465-2012, 2012.
- Wolberg, G. and Alf, I.: An energy-minimization framework for monotonic cubic spline interpolation, *J. Comput. Appl. Math.*, 143, 145–188, 2002.