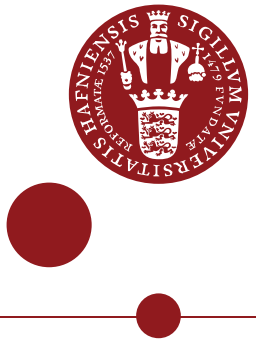
UNIVERSITY OF COPENHAGEN



Assessing the validity of an ice core temperature profile generated from satellite data

A comparison of infrared satellite imagery and ice core temperature profile data for Renland, Greenland

Elloise Fangel-Lloyd

Supervised by Bo Vinther and Jacob Høyer

June 2020

Elloise Fangel-Lloyd
Assessing the validity of an ice core temperature profile generated from satellite data
BSc in Physics, June 2020
Supervisors: Bo Vinther and Jacob Høyer

University of Copenhagen
Faculty of Science
Niels Bohr Institute
Blegdamsvej 17
2100 Copenhagen Ø

Danish Meterological Institute
Centre for Ocean and Ice
Lyngbyvej 100
2100 Copenhagen Ø

Acknowledgements

First of all, many many thanks to my supervisors, Bo and Jacob. Your help and patience is greatly appreciated. I am lucky to have had the opportunity to do such a unique and interesting project. It has been a hugely rewarding experience for me and I thank you very much both for your time.

Many thanks also to my family for their endless support (and proofreading). The extra pairs of eyes are massively appreciated.

Abstract

This work analyses the relationship between RECAP data and a modelled ice core temperature profile created using DMI satellite data, initialised using the step model and the Crank-Nicolson numerical regime as well as the Herron-Langway density model. Contrary to expectation, the cold bias caused by cloud-covered images in the satellite data initially appeared to be insignificant. However, this work suggests that disproportionate data removal due to cloud cover during the cold winter months causes a compensating bias. Additionally, the methodology described above verified that the T_{2M} model contains a warm offset presumably caused by the two-metre height, as expected.

Dette projekt analyserer forholdet mellem RECAP data og en iskerne temperaturprofil fremstillet med DMI satellitdata, initialiseret ved brug af step modellen, Crank og Nicolsons numeriske metode og Herron-Langway densitetsanalyse. Mod forventning var den kolde bias forårsaget af skydække i datasættet tilsyneladende ikke signifikant. Dog foreslår projektet at ujævn datasletning, grundet skydækket under den kolde vinter, producerer en bias som kompenserer. Derudover bruges de ovennævnte metoder til at bekræfte at T_{2M} modellen er præget af et varmt offset, som forventet, formentligt grundet to-meter højden.

Contents

1 Introduction **1**

1.1 Renland and RECAP 1

1.2 Satellite data 2

2 Methods **3**

2.1 Pre-processing 3

2.2 The cloud mask algorithm 3

2.3 Modeling ice flow 4

2.4 Density correction 8

2.5 Statistics 10

3 Results **12**

3.1 Cloud mask analysis 12

3.2 Borehole temperature model 13

4 Discussion **15**

4.1 Cloud mask algorithm 15

4.2 Borehole temperature model 16

5 Conclusion **19**

Bibliography **20**

1 Introduction

1.1 Renland and RECAP

Surface temperature data is used in modelling future climate, and an example of an early climate model is Aristotle's four climate zones, which involves a sphere of fire encircling the planet (Rasmussen, 2010). More recent models use satellite imagery, which provides temperature data with high temporal and spatial resolution, which is particularly relevant for areas such as Greenland, where harsh conditions mean that weather stations are few and far between.

Renland is located on the west coast of Greenland near Scoresbysund, and is of interest for understanding Arctic sea ice conditions during the Glacial and Holocene periods (Johnsen et al, 1992). An ice core provides a measure of how variables change with depth, analogous with time. An ice core is removed using a drill, and a thermometer measures the borehole temperature shortly afterwards. This borehole temperature profile contains information about the geothermal heat from below bedrock as well as the quality of the ice core.

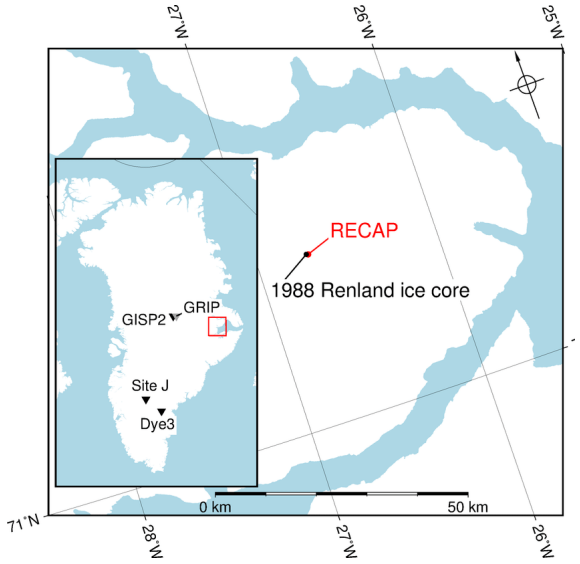


Figure 1.1: Map showing the drill site Renland 1988 and the 2015 RECAP core. Previous drill sites also shown. From Taranczewski et al, 2019.

The original Renland ice core was drilled in 1988, at a length of 300 m (Johnsen, 1992). In 2015, scientists returned to the site (within 2 km; Simonsen et al, 2019) and drilled the RECAP (REnland ice CAP project) core, at a length of 584 m (Wahl, 2018). RECAP is located at 71.30 °N, 26.72 °W (Simonsen et al, 2019).

A previous borehole temperature model for the RECAP core showed that a novel step model accounting for a layer of dead ice was more appropriate than the kink model (Wahl, 2018). Their fit parameters for density modelling were used in this work.

During the drilling process of the RECAP core, it was noted that the drilling liquid was applied only below a certain depth, and so some uncertainty is expected from a "sloshing" effect (Vinther, 2020). This is visible upon inspection of Figure 3.3 and cannot be accurately corrected in this work.

1.2 Satellite data

Borehole temperature profiles provide information on temperature changes with depth but do not contain information on surface temperature over time. Satellite imagery measures surface temperature with high temporal resolution. To compare the two data series, the satellite imagery must first be converted to an effective borehole temperature profile.

The most significant limitation of the satellite data is that cloud cover in the photograph introduces an error to the data (Karlsson and Dybbroe, 2010).

This work also evaluates the T_{2M} model (from the ERA-Interim global reanalysis (Dee et al, 2011) developed by the European Centre for Medium-Range Weather Forecasts), against satellite data. The T_{2M} model generates temperature results two metres above the surface using multiple input parameters that are independent of this satellite and borehole data. T_{2M} is documented to be higher than the measured surface temperature 85% of the time, with the smallest differences occurring during non-melting conditions (Nielsen-Englyst et al, 2019). This work generates a T_{2M} borehole temperature profile in order to examine validity of the T_{2M} result in this application.

2 Methods

2.1 Pre-processing

Code was written using standard Python libraries to sort the 100 gigabytes of satellite data for latitude and longitude, generating data from a 16km^2 grid-square over the RECAP site (Høyer, 2020). This produced a data set of 4 megabytes which was more wieldy to analyse on a personal computer. The sorted data was changed from `netCDF` to `txt` format to improve processing time.

2.2 The cloud mask algorithm

Satellite image quality varies greatly with local weather conditions and cloud cover images must be removed (Karlsson and Dybbroe, 2010). This cloud pre-processing is performed by the Danish Meteorological Institute (DMI), where an algorithm accepts factors including temperature, landscape topography, observed texture and image quality, among others, to determine whether a pixel should be removed (Høyer, 2020). This cloud mask algorithm removes images stochastically, determining the likelihood of an image being misrepresentative.

Arctic surface temperatures are often very close to cloud-top temperatures, and so the selection is not entirely reliable (Karlsson and Dybbroe, 2010). Cloud-top temperatures are likely to still be present in the data, resulting in a bias towards colder temperatures (Høyer, 2020). The magnitude of this bias is not quantified; this work aims to better understand satellite data biases by comparison with the RECAP borehole temperature model.

Examining the number of measurements in a given time period indicates how many images have been removed and gives an indication of data quality. This

work suggests that disproportionate removal of images across different time scales will introduce bias to the data. Measurement days in a year (the number of days in a year for which at least one image is available) as well as monthly measurement fraction (the fraction of a month for which at least one daily measurement is available; this is adjusted for differing lengths of months) are analysed. Renland is north of the Arctic Circle (Simonsen et al, 2019) and so does not experience a twenty-four hour day/night cycle. Therefore, it is not expected that any variation in data removal on a daily scale will be significant. The data selection is poorer in winter (Karlsson and Dybbroe, 2010) and so more data removal is expected in the colder months, causing a warm bias.

2.3 Modeling ice flow

The satellite data is used in an ice flow model in order to produce a model borehole temperature profile, which can then be compared to the RECAP profile.

The earliest ice flow model is the Nye model, which assumes constant vertical strain and constant horizontal velocity throughout the ice sheet (Cuffey and Paterson, 2010). This was followed by the Dansgaard-Johnsen "kink" model, which assumes the horizontal velocity to be constant down to a certain point in the ice sheet, called the "kink height" h. Below this point, horizontal velocity is a function of height above bedrock, surface horizontal velocity, and kink height h (Cuffey and Paterson, 2010).

All ice flow models start with the premise that mass must be conserved, as per the continuity equation.

$$\frac{\delta \rho}{\delta t} + \frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} + \frac{\delta w}{\delta z} = 0 \tag{2.1}$$

Here u denotes velocity in the x direction, v velocity in the y direction, w velocity in the vertical z direction and ρ is density.

It is now assumed that the flow is only two-dimensional, implying that v is very small compared to u. This simplifies derivations, and is valid because

the coordinate axes *x* and *y* can generally be chosen such that the condition is satisfied.

Incompressibility is also assumed, implying constant ρ. This is a valid assumption for ice, but a poor approximation for firn (Cuffey and Paterson, 2010). Precipitation over ice caps often falls as snow, and pressure from overlying load is required before snow is compacted into ice. The between state is called firn, and for shallow parts of an ice core, a density correction term is required. For the purposes of derivation, however, the ice is incompressible. As such Equation 2.2 can be used to calculate vertical velocities.

$$\frac{\delta u}{\delta x} = -\frac{\delta w}{\delta z} \tag{2.2}$$

The horizontal surface velocity is given by Equation 2.3, where *H* is the surface height or thickness of the ice sheet, λ_H the annual accumulation rate of precipitation, H_e the effective height given by equation 2.4 and *x* the distance from the ice divide, which indicates the origin for this coordinate system.

$$u_H = \frac{\lambda_H}{H_e} x \tag{2.3}$$

The ratio of surface to bedrock horizontal velocities, or bottom sliding coefficient f_b, is introduced, and is present due to the frictional forces exerted by the bedrock. The effective height is defined as the height that the ice sheet would be, if the velocities were constant with height.

$$H_e = H + \frac{h}{2}(f_b - 1) \tag{2.4}$$

A step-function limited by the kink height expresses horizontal velocity as a function of height above bedrock.

$$u(z) = \begin{cases} u_H & h \leq z \leq H \\ u_H(f_b - \frac{z}{h}(1 - f_b)) & 0 \leq z \leq h \end{cases} \tag{2.5}$$

Equation 2.2 can be used to find the vertical velocities by integrating within the given limits.

In this work, the step model will be used, as it has proved a significantly better fit to the RECAP ice core data (Wahl, 2018). This replaces the kink height with a step height h_s, defined by the kink height minus the height of an assumed

layer of dead ice, which has zero velocity in all directions. The horizontal (x-direction) velocities as a function of height are then given by

$$u(z) = \begin{cases} u_H & h_s \leq z \leq H \\ u_H f_b & 0 \leq z \leq h_s \end{cases} \tag{2.6}$$

which defines the vertical velocities below. u_H is as previous and H_e is now defined by substituting h for h_s (Wahl, 2018).

$$w(z) = \begin{cases} -\frac{\lambda_H}{H_e}(z - h_s(1 - f_b)) & h_s \leq z \leq H \\ -\frac{f_b \lambda_H}{H_e} z & 0 \leq z \leq h_s \end{cases} \tag{2.7}$$

2.3.1 Steady-state model

In order to accurately model an ice core, a steady-state model must be used as an initial condition. Additional corrections are then applied to it to simulate the non-steady-state case. The model is numerically initialised using the Crank-Nicolson regime and the heat transfer equation (Crank and Nicolson, 1996).

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial z^2} + w(z)\frac{\partial T}{\partial z} + u(x)\frac{\partial T}{\partial x} \tag{2.8}$$

The first term on the right-hand side is the conduction term. It involves κ, the thermal diffusivity of ice, which depends on the density, specific heat capacity and thermal conductivity of ice. The second term describes advection in the vertical dimension and the third advection in the horizontal *x* dimension. Temperature gradients are far more significant in the vertical than in the horizontal, so the horizontal advection is neglected.

In order to numerically solve Equation 2.8, it is discretized. This is done by introducing indices i and j, which refer to steps in time Δt and steps in height Δz, respectively. The discrete form of the heat transfer equation is then (Crank and Nicolson, 1996)

$$\begin{aligned} \frac{T_{i+1,j} - T_{i,j}}{\Delta t} = \frac{\kappa}{2}\left(\frac{T_{i+1,j+1} - 2T_{i+1,j} + T_{i+1,j-1}}{2\Delta z^2} + \frac{T_{i,j+1} - 2T_{i,j} + T_{i,j-1}}{\Delta z^2}\right) \\ -\frac{w(z)}{2}\left(\frac{T_{i+1,j+1} - T_{i+1,j-1}}{2\Delta z} + \frac{T_{i,j+1} - T_{i,j-1}}{2\Delta z}\right) \end{aligned}$$

this rearranges to

$$T_{i+1,j} - T_{i,j} = \frac{\kappa \Delta t}{2\Delta z^2}(T_{i+1,j+1} - 2T_{i+1,j} + T_{i+1,j-1} + T_{i,j+1} - 2T_{i,j} + T_{i,j-1}) \\ -\frac{w(z)\Delta t}{4\Delta z}(T_{i+1,j+1} - T_{i+1,j-1} + T_{i,j+1} - T_{i,j-1})$$

Diffusion and convection terms can then be defined, respectively

$$s_j = \frac{\kappa \Delta t}{2\Delta z^2} \qquad\qquad r_j = \frac{w(z)\Delta t}{4\Delta z}$$

To numerically analyse the model, all current timesteps i must be sorted to one side of the equation and all next steps $i+1$ to the other. The coefficients s_j and r_j can then be applied to simplify.

$$(r_j - s_j)T_{i+1,j+1} + (1 + 2s_j)T_{i+1,j} + (-r_j - s_j)T_{i+1,j-1} = \\ (s_j - r_j)T_{i,j+1} + (1 - 2s_j)T_{i,j} + (s_j + r_j)T_{i,j-1} \tag{2.9}$$

This is rewritten to matrix notation.

$$\begin{bmatrix} b_1 & c_1 & 0 & \dots & 0 \\ a_2 & b_2 & c_2 & \ddots & \vdots \\ 0 & a_3 & b_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & c_{n-1} \\ 0 & \dots & \dots & a_n & b_n \end{bmatrix} \begin{bmatrix} T_{i+1,1} \\ T_{i+1,2} \\ T_{i+1,3} \\ \vdots \\ T_{i+1,n} \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_n \end{bmatrix}$$

Here, the f elements are the right-hand side of equation 2.9, representing the temperature at the current timestep. The vector T contains the temperatures for all Δz at the next timestep, and it is calculated by matrix inversion.

$$\mathbf{M}T = f \longrightarrow T = \mathbf{M}^T f$$

The orientation of the matrices is such that the first row corresponds to bedrock and the nth row to the surface temperature.

In the steady-state model, the only nonzero elements in the f-vector are f_1 and f_n. f_1 corresponds to the temperature gradient at bedrock, or the geothermal

heat flux from the rock into the ice, defined by Equation 2.10, with K the thermal diffusivity.

$$f_1 = \frac{Q_{geo}\Delta z}{K} \tag{2.10}$$

To generate a steady-state ice core model, a measurement of surface temperature is used as input f_n. All ice core parameters are assumed to be the same as used to model the RECAP ice core. This input parameter is -19.2 C, as this was the measured ambient temperature when the core was drilled (Vinther, 2020). The matrix elements defined below are the same for both steady- and non-steady-state models.

$$b_1 = -1 \qquad c_1 = 1 \qquad a_n = 0 \qquad b_n = 1$$

2.3.2 Non-steady-state model

The steady-state-model is used as an initial condition to define a non-steady-state model. In the non-steady-state model, the f values are all nonzero.

$$f_j = (s_j - r_j)T_{i,j+1} + (1 - 2s_j)T_{i,j} + (s_j + r_j)T_{i,j-1} \tag{2.11}$$

To generate a borehole temperature profile based on satellite data, the matrix must be calculated and inverted once for each satellite data point. f_1 is defined as previously. f_n is again surface temperature, here measured by satellite. To calculate a satellite timestep, the temperature values of the preceeding timestep are used, as per Equation 2.11. The first timestep uses the steady-state model to initialise, and after the analysis of each satellite timestep, the array is updated. The final timestep analysis then contains the fully corrected non-steady-state model.

2.4 Density correction

Snow is an extremely porous medium, consisting of an ice matrix with corridors of empty space. As long as these spaces exist (until the firn is sufficiently compressed), gas can diffuse from the surface through the ice core (Gkinis, 2020). As many of the above model parameters are density-dependent, density

must be corrected in the upper layers. Three stages of densification are defined (Herron and Langway, 1980).

1. The **initial** stage (ρ < 550 kg m^{-3})
   Densification takes place very quickly. The critical density is defined by densification taking place more slowly. $\rho_{crit} = 550$ kg m^{-3}.

2. The **second** stage (550 $< \rho <$ 820 kg m^{-3})
   Densification is now much slower. As the density approaches the close-off density, $\rho_{co} = 820$ kg m^{-3}, the air pockets close and become bubbles rather than channels.

3. The **close-off** stage ($820 < \rho <$ 917 kg m^{-3})
   90% of the entrapped air has ascended back to the surface (Gkinis, 2020). Existing bubbles continue to compress until the density of ice $\rho_{ice} = 917$ kg m^{-3} is reached. Density is constant at ρ_{ice} below this.

The Herron-Langway model is used to correct density. Herron-Langway analysis assumes that there is a linear relationship between the change in stress as a result of mass of overlying firn and proportional change in air spaces.

$$\frac{\mathbf{d}\rho}{\rho_{ice} - \rho(d)} = k\rho \ \mathbf{d}d \tag{2.12}$$

Here k is a constant and d depth. This equation is manipulated to find a two-stage model with expressions for the initial and second stages (Herron and Langway, 1980).

To correct for density, firn must be accounted for in Equation 2.9. The thermal convection term r_j is modified and a term relating to firn compaction is introduced, as below (Wahl, 2018).

$$r_j = \frac{\Delta t}{4\Delta z}\left(w(z) + 0.0057\kappa D - \frac{\partial \rho}{\partial z}\left(\frac{\kappa}{\rho} - \frac{\partial \kappa}{\partial \rho}\right)\right) \tag{2.13}$$

where

$$D = \frac{T_{i,j+1} - T_{i,j-1}}{2\Delta z}$$

Equation 2.13 now accounts for density and temperature dependence of the various parameters. The firn compaction term is introduced to the f-vector.

$$f_j = (s_j - r_j)T_{i,j+1} + (1 - 2s_j)T_{i,j} + (s_j + r_j)T_{i,j-1} + L\frac{\Lambda g}{\rho^3 c}\frac{\partial \rho}{\partial z} \tag{2.14}$$

Where L is the overlying load of snow (kg), Λ the annual accumulation rate in kg m^{-1} year^{-1}, g the acceleration due to gravity and c the specific heat capacity of ice, given by

$$c = 152.5 + 7.122T$$

Equation 2.14 represents a non-steady-state, density-corrected borehole temperature profile that takes satellite data as input.

2.5 Statistics

In order to compare two data series, the Pearson chi-square (least squares) test as well as the Wald-Wolfowitz Runs test are used.

The reduced chi-square is found by dividing the chi-square by the number of degrees of freedom. This corrects for large datasets and effectively normalises (Barlow, 1989). The reduced chi-square statistic, χ^2_{red}, should be as close to 1 as possible. $\chi^2_{red} \gg 1$ indicates bad model fit. $\chi^2_{red} > 1$ indicates an error variance, or spread, that has not been accounted for, whereas $\chi^2_{red} < 1$ indicates over-estimation of spread (Barlow, 1989). Chi-square calculates the difference between two data series, and does not discern between difference due to spread and difference due to an incorrect model. To evaluate the choice of model, a Runs test is used instead.

The Wald-Wolfowitz Runs test produces n points when used to compare two data series, each n long. Let one data series be A, and the other B. (Choice does not matter as long as consistency applies.) Each point in the Runs then consists of either a "+" if A is larger than B at that specific point, and a "-" if the reverse is true. The result is a vector consisting of "+" and "-" (Vidakovic, 2011).

Following this, the number of "+" runs n_+ is calculated by simple sum. The overall Runs statistic for the data series R is then

$$R = \frac{n_+}{\frac{n}{2}}$$

Ideally, the number of "+" would exactly equal the number of "-", producing a Runs statistic of 1. This indicates very good general model fit, but does not test for spread.

This work will use reduced chi-square and Runs statistics to complement each other. The former quantifies spread and the latter model fit. This allows for more detailed and quantitative analysis of systematic errors as well as more thorough discussion of model validity than by only using one of the two (Vidakovic, 2011).

3 Results

3.1 Cloud mask analysis

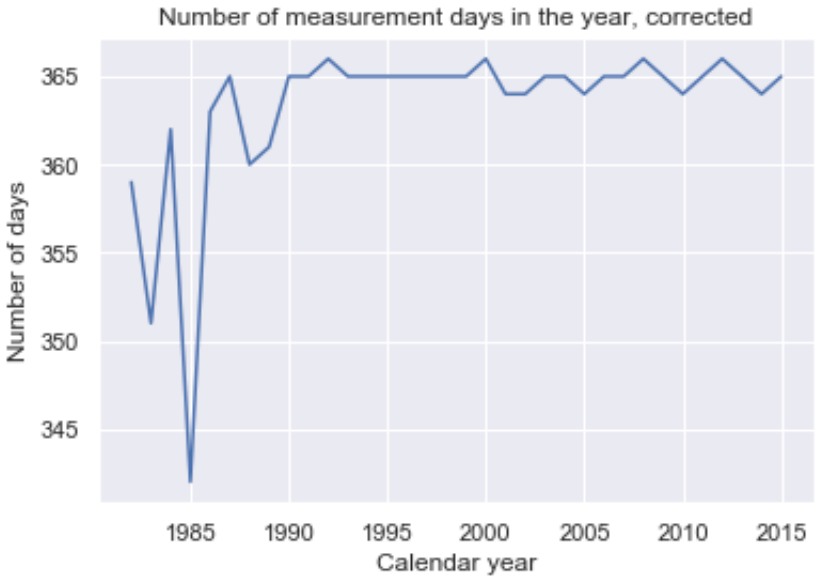


Figure 3.1: Number of days in the year for which at least one measurement is available, corrected for 1997 anomaly. The data is Boolean so there is no error.

An anomaly was present in the 1997 data, where only 201 measurement days were available. As this presents a significant error in the dataset, the 1997 data were removed and replaced with an interpolation from 1996-98, as shown on Figure 3.1.

Figure 3.2 shows monthly measurement fraction with errorbars $\pm 1\sigma$. It is worth noting that the errorbars extend into the unphysical. Figure 3.2 shows measurement fraction of a month; by definition it cannot be larger than 1. The errorbars extend above 1 because they are $\pm 1\sigma$; naturally, a realistic error range would cap at 1.

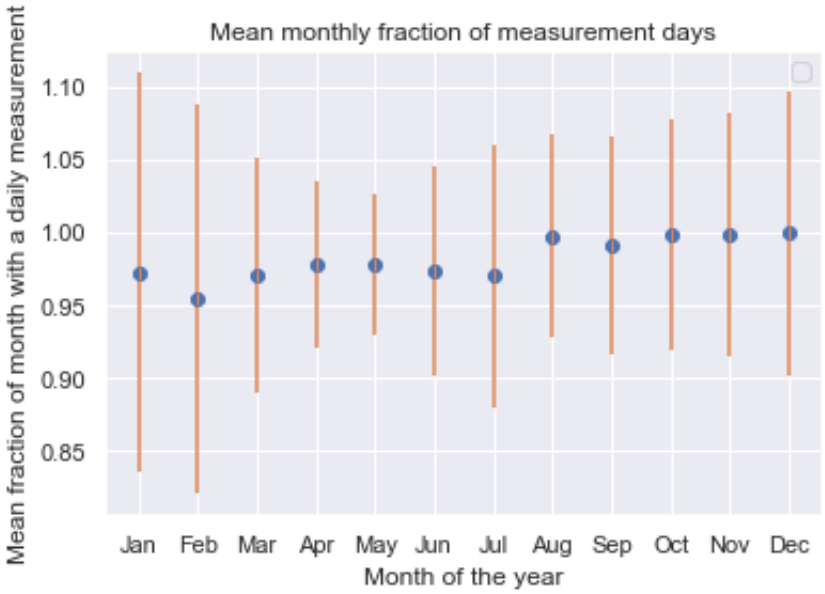


Figure 3.2: Monthly measurement fraction averaged over all available years. Y-axis extended for viewing detail. Errorbars representing standard deviation, 1σ.

3.2 Borehole temperature model

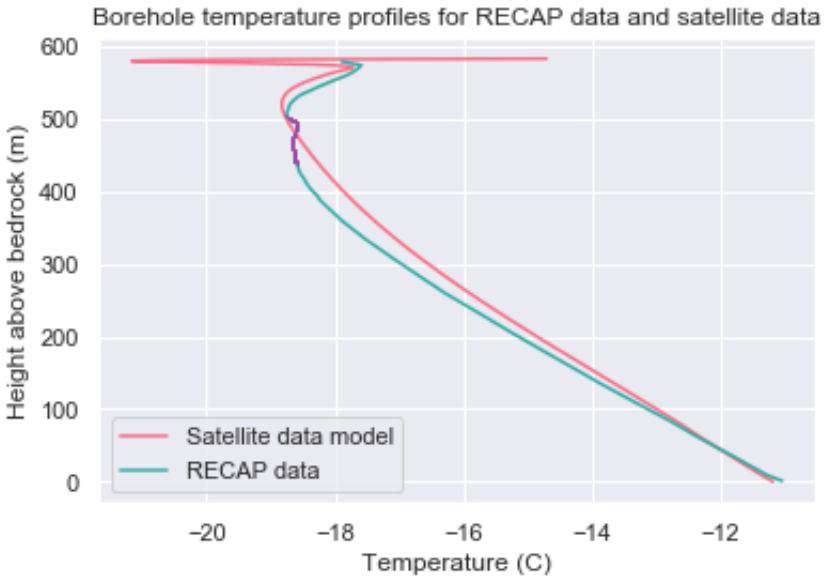


Figure 3.3: Non-steady-state density-corrected borehole temperature model based on satellite surface temperature data. For clarity, errorbars (RECAP only) and data points are not shown. RECAP data affected by drilling liquid issues shown in purple.

Figure 3.3 shows a borehole temperature profile for the satellite data model as described above, with RECAP data. The aforementioned experimental issues

with drill liquid are shown in purple. RECAP errorbars are not shown as they are very small; the stated instrument uncertainty is 0.025%.

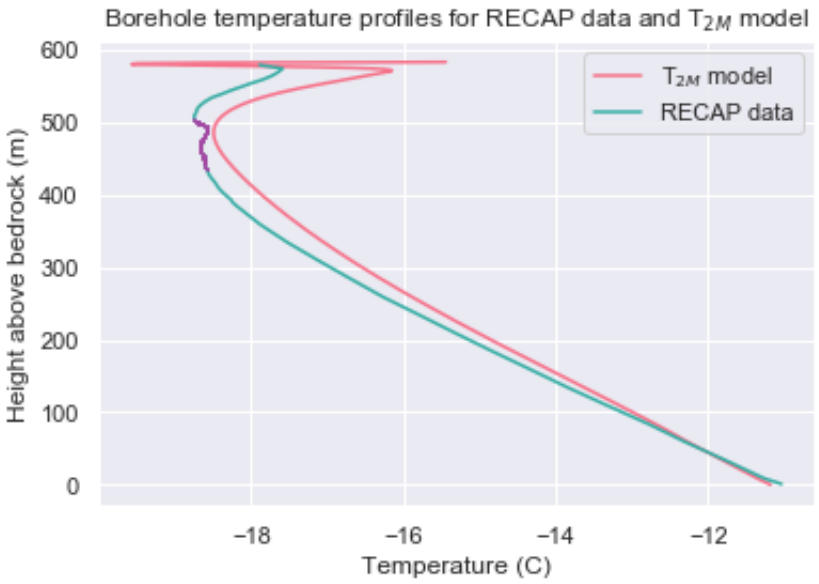


Figure 3.4: Non-steady-state density-corrected borehole temperature model based on T_{2M} data. For clarity, errorbars (RECAP only) and data points are not shown. RECAP data affected by drilling liquid issues shown in purple.

Figure 3.4 shows a borehole temperature profile for the T_{2M} model data, with RECAP data.

	Satellite data	T_{2M}
R	1.71	1.93
χ^2_{red}	0.813	0.740

Table 3.1: Table showing statistical parameters for the borehole temperature model. Runs statistic and reduced chi-square shown.

To analyse model fit quality, a tuning was performed. This allowed surface temperature to vary, as well as an offset which was added to the final satellite data profile, essentially moving it along the x axis on Figure 3.3. Both of these were permitted to vary between -0.5 and 0.5 degrees, and statistical parameters were calculated for each. It was found that there was no improvement on either the runs parameter or the chi-square from tuning these variables. Tuning results are given in Appendix B.

4 Discussion

4.1 Cloud mask algorithm

4.1.1 Annually

Figure 3.1 shows the corrected number of measurement days, where an interpolation has been performed to replace the 1997 data. The ideal number of measurement days would be 365 with 366 in four-year increments; this was not anticipated, as cloud cover is prevalent over Greenland (Karlsson and Dybbroe, 2010). Initially the number of measurement days fluctuates drastically, but after 1990 it approaches a full year.

The pre-1990 variation in measurement days is not necessarily indicative of any specific patterns in cloud cover. It appears to be the result of substandard satellite technology, as it ceases to fluctuate so dramatically in later years (Høyer, 2020). An unknown instrument issue is assumed to be the cause, such as satellite replacement with improved technology.

Using a mean and a standard deviation ($\pm 1\sigma$), the number of measurement days (averaged over the whole period analysed) can be expressed as

$$363.3 \ \pm \ 4.6 \text{ days}$$

which indicates very good temporal resolution on the annual scale. Realistically, equipment limitations and weather conditions mean that improvements on this are not to be expected.

It is interesting to note that the bias due to missing data is largely due to pre-1990 measurements. Were this work to be repeated, removing pre-1990 data would increase the mean number of measurement days and minimise

this bias. Correcting for the differences in errors pre- and post-1990 is seen as beyond the scope of this work, but could be a valuable future analysis.

4.1.2 Monthly

Figure 3.2 shows the mean monthly fraction of measurement days, which should be close to 1. Again, using a mean and standard deviation, the measurement fraction can be expressed as

$$0.982 \pm 0.0854$$

It is clear that data has been removed, but a removal of data will not generate bias if it is randomly sampled. The uncertainty is significant - $\pm$ 8.7% - indicating that there is spread in data removal between months, and so there is a non-random sampling and a bias is introduced to data.

The errorbars on Figure 3.2 are significantly larger in the winter months than the rest of the year, indicating that more data is removed in these months. This confirms the presence of a warm bias.

There is the issue of the cold bias caused by remaining cloud cover in the data. This is unquantifiable in magnitude, but it is possible that the warm bias arising from winter data being removed is to some extent "cancelling out" the cold bias. Visual inspection of Figure 3.3 cannot conclusively identify one bias or the other, nor can the statistical parameters. As such, it is plausible that either the warm bias is extremely small (and the cold bias is also negligible), or the biases cancel.

4.2 Borehole temperature model

Figure 3.3 shows the borehole temperature model plotted with RECAP data. Visual inspection shows dramatic behaviour near the surface, which arises for experimental reasons (Vinther, 2020). The swing towards colder regions is the previous winter's precipitation whose low temperature is preserved by the layers above it. The sharp swing towards warmer temperatures is due to the relatively high surface temperature in the early summer. This is not observed

on the RECAP data because temperature measurements begin at 5m below the surface (Vinther, 2020). As such, it is not possible to accurately quantify this work's satellite data model at the surface.

Comparison of two data series benefits datasets of the same length, so rather than interpolate the RECAP data, a downsampling was performed on the borehole temperature profile.

In the Runs test used in this work $R = 1.71$ was determined for the satellite borehole model. This is seen as an acceptable runs parameter, given that the experimental "wobble", associated with the drilling liquid issue on the RECAP data, causes R to be larger.

The R-value for the T_{2M} model is 1.93. This is markedly larger than for the satellite borehole model. The value $R = 2$ indicates a model that is warmer than the data in *all* points, i.e. a consistent warm bias, as per Section 2.5. Visual inspection of Figure 3.4 confirms a a clear warm bias from 50 m and upwards. This is as expected; as discussed previously, the T_{2M} is consistently warmer than the satellite data. This Runs analysis indicates that the satellite borehole model correlates with RECAP data much better than the T_{2M} borehole model.

The χ^2_{red} of the satellite borehole model (Table 3.1) indicates that the spread is of an acceptable magnitude. Visual inspection of Figures 3.3 and 3.4 suggests that the χ^2_{red} value may be significantly closer to one if the data near the very surface were ignored. The reduced chi-square for the T_{2M} data indicates a significantly worse fit, and confirms that the satellite data correlates better with RECAP data than the T_{2M} borehole model. Overall, as both $\chi^2_{red} < 1$, an over-estimation of spread may have taken place (Barlow, 1989).

The tuning performed indicated that no change in either surface temperature or offset improved the model fit. This indicates correct choice of model as well as accurate initial parameters.

As previously discussed, a warm bias (Section 4.1.2) is present in the data, with an uncertainty on monthly measurement fraction of 8.7%. It is also noted that the performed tuning found no improvement. This leads the author to

conclude that either the warm bias is very small, or both aforementioned warm and cold biases are present, but to some degree compensate each other.

Having only 56 data points places some limits on statistical analysis. This is a limitation of the RECAP data rather than the model; however, the RECAP data fits the kink-model well (Wahl, 2018) and is overall good-quality data, drilling liquid issues aside. Statistical analysis of the effect of increasing the number of data points may be relevant further work.

As the cold bias cannot be directly quantified, this work is suggesting that a more complete evaluation requires the removal of the possible warm bias. Further work could include weighting to compensate for missing data due to cloud cover.

Conclusion

5

Analysis of the cloud mask algorithm found that prior to 1990, the number of measurement days in a year fluctuated but became more stable after 1990, with a more acceptable range of 363.3 $\pm$ 4.6 days.

On a monthly scale, it was expected that more data would be removed in winter, which was the case. Overall a measurement fraction of 0.982 $\pm$ 0.0854 was found. This indicates that a small spread is present and as such a warm bias due to removal of data is present. Methods to compensate have been discussed.

Borehole temperature profiles were generated for both the satellite data and the T_{2M} model. Statistical analysis using both Runs test and χ^2_{red} found that the satellite data profile was better suited to RECAP than T_{2M}. This is not unexpected, as T_{2M} is from two metres above the surface.

The performed tuning found that applying an offset and changes to surface temperature did not produce improved statistical parameters for Runs or chi-square. This taken in conjunction with the proven presence of a warm bias, supports the conclusion that either the anticipated cold bias and this warm bias are compensating each other, or they are both small in magnitude.

Further work might contain a weighting or interpolation of satellite data, to remove the warm bias from the data and more conclusively quantify the cold bias. Alternatively, further work could address the removal of pre-1990 data or use the RECAP density data to calibrate the Herron-Langway model.

Bibliography

[1] Rasmussen, E., 2010. *Vejret Gennem 5000 År*. Aarhus: Aarhus Universitetsforlag.

[2] Johnsen et al, 1992. *A "deep" ice core from east Greenland*, MoG Geoscience, vol. 29.

[3] Taranczewski, Tetsuro et al. *10,000 years of melt history of the 2015 Renland ice core, East Greenland*. The Cryosphere Discussions (2019): 1-16.

[4] Simonsen et al. *East Greenland ice core dust record reveals timing of Greenland ice sheet advance and retreat*. Nature Communications 10, 4494 (2019).

[5] Wahl, S., 2018. *Holocene Climate And Ice Flow Of The Renland Ice Flow*. Master Thesis. University of Copenhagen and Universitat Hohenheim.

[6] Vinther, B., 2020. *The RECAP Drilling*. Personal communication.

[7] Karlsson, K. and Dybbroe, A., 2010. *Evaluation of Arctic cloud products from the EUMETSAT Climate Monitoring Satellite Application Facility based on CALIPSO-CALIOP observations*. Atmospheric Chemistry and Physics, 10(4), pp.1789-1807.

[8] Dee et al, 2011. *The ERA-Interim reanalysis: configuration and performance of the data assimilation system*. Quarterly Journal of the Royal Meteorological Society, vol. 137, no. 656, pp. 553 - 597.

[9] Høyer, J., 2020. *Cloud Mask Algorithm*. Personal communication.

[10] Nielsen-Englyst et al, 2019. *In situ observed relationships between snow and ice surface skin temperatures and 2 m air temperatures in the Arctic*. Cryosphere. 13. 1005-1024.

[11] Cuffey, K. and Paterson, W., 2010. *The Physics Of Glaciers*. 4th ed. Burlington: Elsevier Science.

[12] Crank, J. and Nicolson, P., 1996. *A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type*. Advances in Computational Mathematics, 6(1), pp.207-226.

[13] Gkinis, 2020. *Densification of polar snow*, lecture notes. Ice Core Glaciology NFYK14025U, University of Copenhagen. Delivered 24 November 2019.

[14] Herron, M. and Langway, C., 1980. *Firn Densification: An Empirical Model*. Journal of Glaciology, 25(93), pp.373-385.

[15] Barlow, R., 1989. *Statistics*. 3rd ed. Chichester: Wiley.

[16] Vidakovic, B., 2011. *Distribution-Free Tests. In: Statistics for Bioengineering Sciences*. Springer Texts in Statistics.

Appendix A

Below is a short Python program which generates a borehole temperature profile from satellite data.

```
import numpy as np
import matplotlib.pyplot as plt

#importing satellite data
monthly = np.loadtxt('corrected.txt', delimiter=',')



##DEFINING PARAMETERS


H = 584-48 #minus for dead ice
h = 217-48
fb = 0.29
He = H - (h*(1-fb))
lam = 0.33/(3600*24*365) #accumulation rate originally in m/year
Lam = lam*917
Q = 48*10**-3 #48 mW geothermal heat
t = 3600*24*30 #time step, one month
delz = 4 #spatial step
n = int(584/delz) #define number of spatial steps
num_points = 403 #number of time steps/model runs
g  = 9.81


z = np.zeros([n,1]) #empty array for z
w = np.zeros([n,1]) #empty array for w
coeffs = np.zeros([n,n]) #empty, defining the large matrix
funcs = np.zeros([n,1]) #empty, defining the f vector

T_vek = np.loadtxt('temps.txt')+273.15 #import the steady state
    initial conditions, convert to K
T_list = [ [] for i in range(num_points+1) ]
T_list[0] = T_vek #first element of T = steadystate guess

##VELOCITIES
```

```
for i in range(0,n): #define z, list of heights, spaced by delta z
    z[i] = i*delz

#defining vertical velocity
for i in range(0,n-1):
    if z[i] > h+48:
        w[i] = -(lam/He) * ((z[i]-48) - (h*(1-fb)))
    if 48 < z[i] < h+48:
        w[i] = -fb*lam*(z[i]-48)/He
    if 0 < z[i] < 48: #if dead ice
        w[i] = 0
w[-1] = -lam


##DENSITY FUNCTION

d = z[::-1] #finds z reversed, for depths rather than heights
d_crit = 7.97 #critical depth
d_co = 57.02 #closeoff depth
rho_0 = 0.391084 #initial density
rho_crit = 0.497782 #critical density
rho_co = 0.820 #close-off density
lh = 0.42 #accumulation rate
Ts = 253.95 #surface temp
R = 8.314 #gas constant
K0 = 11 * np.exp(-10160/(R*Ts)) #Arrhenius type numbers
K1 = 575 * np.exp(-21400/(R*Ts))
rho_ice = 0.917 #density of ice


rho = [] #defining empty list for density
#rho as function of depth

Z1 = np.zeros([n])#empty array for Z1 func
Z0 = np.zeros([n]) #empty array for Z0 func


#defining density
for i in range(0,n):

    if d[i] == 0: #density at surface
        rho.append(rho_0)

    if 1 < d[i] < d_crit: #critical density and above
```

```
        Z0[i] = np.exp((rho_ice*K0*d[i]) + np.log(rho_0/(rho_ice -
    rho_0)))
        rho.append((rho_ice * Z0[i])/(1 + Z0[i]))

    if d_crit < d[i] < d_co: #between critical and closeoff density
        Z1[i] = np.exp((rho_ice*K1*(d[i]-d_crit)*(lh**-0.5)) + np.
    log(0.55/(rho_ice - 0.55)))
        rho.append((rho_ice * Z1[i])/(1 + Z1[i]))

    if d_co < d[i]:  #below close off depth, constant density
        rho.append(rho_ice)

#defining density profile

##DENSITY DEPENDENT PARAMS

rho_z = rho[::-1] #finding rho as function of z

L = [] #defining L parameter
for i in range(0,n):
    L.append(z[i]*rho_z[i])


#defining thermal conductivity of ice
K_ice = []
for i in range(0,n):
    K_ice.append(9.828 * np.exp(-0.0057*T_vek[i]))

#defining specific heat capacity
    #using steady-state guess
c = []
for i in range(0,n):
    c.append(152.5 + (7.122*T_vek[i]))

#thermal conductivity of firn
K_firn = []
for i in range(0,n):
    K_firn.append(K_ice[i] * ((rho[i]/rho_ice)**(1-(rho[i]/(2*
    rho_ice)))))


#thermal diffusivity
kappa_firn = []
for i in range(0,n):
```

```
    kappa_firn.append(K_ice[i]/(c[i]*rho_ice*1000) * ((rho[i]/
    rho_ice)**(1-(rho[i]/(2*rho_ice)))))

#drho/dz
rho_func = []
for i in range(0,n-1):
    rho_func.append((rho_z[i+1]-rho_z[i])/delz)

#dkappa/drho
kappa_func = []
for i in range(0,n):
        kappa_func.append(kappa_firn[i]*((1/(rho_z[i]*1000)) - \
            (1/(2*rho_ice*1000))) * (1+np.log((rho_z[i])/(rho_ice))
    ))


##MODELING

#define list of model runs in Celsius for plotting
temps = [ [] for i in range(num_points+1) ]
temps[0] = T_list[0] - 273.15

#defining the matrix

for j in range(1,num_points): #looping over model runs
    #begins at 1, as the first element in T_list is already defined


    for i in range(0,n): #looping over depth steps


        if i == 0: #first row

            coeffs[i,i] = -1 #b1
            coeffs[i,i+1] = 1 #c1

            funcs[i] = -Q*delz/K_ice[-1] #f1


        if i == n-1: #last row

            coeffs[i,i] = 1 #bn
            coeffs[i, i-1] = 0 #an

            funcs[i] = monthly[j] #fn; accepts satellite data
```

```
    elif i != 0 and i != n-1:

        D = [(T_list[j-1][i+1] - T_list[j-1][i-1])/2*delz \
             for i in range(n-1)] #defining new list D

       #defining rj and sj coefficients

        rj = t/(4*delz) * (w[i] + 0.0057 * D[i] * kappa_firn[i]
 \
        - (rho_func[i]*((kappa_firn[i]/rho[i]) + kappa_func[i])
))



        sj = (kappa_firn[i] * t)/(2*(delz**2))
        #defining sj

        coeffs[i,i] = 1+(2*sj) #bj
        coeffs[i,i-1] = -rj - sj #aj
        coeffs[i,i+1] = rj - sj #cj

        funcs[i] = (sj - rj)*T_list[j-1][i+1] + \
        (1-(2*sj))*T_list[j-1][i] + (sj + \
        rj)*T_list[j-1][i-1] + \
        (L[i] * (Lam*g/(c[i]*(rho[i]*1000)**3)) * (rho_func[i])
 )#fj


 T_list[j] = np.dot(np.linalg.inv(coeffs), funcs) #inverts the
matrix
 temps[j] = T_list[j][:] - 273.15 #converts all to Celsius, for
plotting
```

Appendix B

Table 5.1 shows reduced chi-square values (no units) calculated for the performed tuning. Along the top, input surface temperature is varied. On the left up to down, offset temperature is varied.

	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5
-0.5	0.513	0.538	0.564	0.591	0.619	0.649	0.679	0.711	0.735	0.748	0.760
-0.4	0.538	0.541	0.552	0.563	0.564	0.576	0.581	0.592	0.603	0.616	0.637
-0.3	0.564	0.591	0.619	0.649	0.679	0.710	0.684	0.671	0.664	0.652	0.631
-0.2	0.613	0.624	0.644	0.668	0.679	0.692	0.712	0.729	0.738	0.769	0.684
-0.1	0.619	0.648	0.679	0.704	0.735	0.776	0.749	0.710	0.689	0.652	0.643
0	0.684	0.703	0.736	0.765	0.797	0.813	0.789	0.775	0.741	0.712	0.689
0.1	0.645	0.678	0.704	0.735	0.773	0.798	0.762	0.728	0.690	0.672	0.631
0.2	0.600	0.628	0.657	0.683	0.704	0.739	0.712	0.695	0.662	0.631	0.604
0.3	0.578	0.599	0.627	0.658	0.683	0.702	0.691	0.654	0.634	0.602	0.568
0.4	0.529	0.571	0.598	0.621	0.657	0.689	0.662	0.619	0.589	0.561	0.531
0.5	0.501	0.532	0.579	0.614	0.657	0.694	0.679	0.649	0.612	0.587	0.554

Table 5.1: Reduced chi-square fit parameters. Top left to right is variation in input surface temperature; up to down is variation in offset.