Impact of the Ice Strength Formulation on the Performance of a Sea Ice Thickness Distribution Model in the Arctic

Mischa Ungermann[1], L. Bruno Tremblay[2], Torge Martin[3] and Martin Losch[1]

Corresponding author: M. Ungermann, Alfred Wegener Institute, Bussestr 24, 27570 Bremerhaven, Germany (mischa.ungermann@awi.de)

[1]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany.

[2]Department of Atmospheric and Oceanic Sciences, McGill University, Montréal, Quebec, Canada.

[3]GEOMAR, Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

Key Points.

- An ice thickness distribution (ITD) model can significantly improve the fit to satellite observations
- The simple ice strength parameterization of *Hibler* [1979] leads to smaller model errors than the one of *Rothrock* [1975]
- The ice strength following *Rothrock* [1975] strongly depends on the number of thickness classes

Abstract. The impact of a subgrid-scale ice thickness distribution (ITD) and two standard ice strength formulations on simulated Arctic sea ice climate is investigated. To this end different model configurations with and without an ITD were tuned by minimizing the weighted mean error between the simulated and observed sea ice concentration, thickness and drift speed with an semi-automatic parameter optimization routine. The standard ITD and ice strength parameterization lead to larger errors when compared to the simple single-category model with an ice strength parameterization based on the mean ice thickness. Interestingly, the simpler ice strength formulation, which depends linearly on the mean ice thickness, also reduces the model-observation error when using an ITD. For the ice strength parameterization that makes use of the ITD, the effective ice strength depends strongly on the number of thickness categories, so that introducing more categories can lead to overall thicker ice that is more easily deformed.

1. Introduction

Reliable sea ice models are an essential ingredient of climate models, but also of accurate sea ice forecasts that are required by the increasing shipping activities in the Arctic. The requirement of accuracy, together with advances in computing power, has led to an increase in sea ice model complexity over the last decades. With the rising amount of available observational data of Arctic sea ice, many new physical processes have been included in additional model parameterizations [*Hunke et al.*, 2011]. For the development of future model systems a thorough scrutiny of each component of a sea ice model as well as its interaction with other components seems necessary [e.g. *Hunke*, 2014].

One of the most commonly used parameterizations in current sea ice models employs a subgrid-scale ice thickness distribution (ITD) to describe the ice thickness in each grid cell. Most implementations today are based on *Thorndike et al.* [1975]. There are two main reasons that motivated this parameterization: First, the conductive heat flux through sea ice is dominated by the contributions of thin ice and open water, even if they cover only a small fraction of the total area. Second, most of the ice deformation processes, especially of a thicker and stronger pack, are ridging of the thinner ice fraction and shearing along leads (also characterized by thin or no ice). Hence, an ITD is used in many sea ice models and many new parameterizations — such as an ice enthalpy distribution [*Zhang and Rothrock*, 2001] or an anisotropic rheology of discrete failure regimes [*Wilchinsky and Feltham*, 2012] — are based on an ITD model. Although ITD models seem to be well established, many questions about the exact mechanics of the involved processes and about the ITD's impact on model simulations remain.

Already when the ITD parameterization originally was developed, two main problems were identified that are still the biggest sources of uncertainty today: (1) the redistribution of ice between different ice thickness categories by ridging processes [*Thorndike et al.*, 1975] and (2) the assumption that the deformation energy is either lost to friction or converted to potential energy as ice floes ridge and raft [*Rothrock*, 1975]. Both *Thorndike et al.* [1975] and *Rothrock* [1975] make assumptions about the mechanical processes that govern sea ice ridge formation, but *Pritchard* [1981] already showed that they were missing important parts of the energy balance. At the time there were only a few observations of thickness and ridge profiles available [see e.g. *Parmerter and Coon*, 1972, and references therein], and dynamical modeling studies provided the most reliable understanding of ridging processes [*Parmerter and Coon*, 1973]. The amount of available data has increased since. After discrete element models of the ridging process [*Hopkins*, 1998], laboratory experiments of ridging [*Tuhkuri*, 2002], and in-situ measurements of stresses in ice floes [*Tucker and Perovich*, 1992; *Richter-Menge and Elder*, 1998], the analysis of ridging properties is still an important field of ongoing research. Methods range from evaluating airborne observations [*Herzfeld et al.*, 2015] and basin-wide process-oriented model simulations [*Hopkins and Thorndike*, 2006] to the analysis of conceptual models [*Godlovitch et al.*, 2011]. A common notion is that the details of the physical processes during ridging and their large-scale statistical properties, that is, the key features in shaping an ITD and determining the amount of energy necessary for deformation, are still not sufficiently well understood.

To evaluate an ITD model in view of uncertain theory, one of the first approaches was to compare the results to observed ice thickness. Such assessments are impeded by the sparsity of observational data for ice thickness. Still, *Thorndike et al.* [1975] could successfully simulate thickness distributions with a column ITD model that were similar to upward looking sonar measurements from submarines sailing under the Arctic sea ice. *Bitz et al.* [2001] reproduced this result in their global coupled model against a much larger set of similar upward looking sonar data. In spite of this partial success, high uncertainties remain in ice thickness data both from models and observations [*Schweiger et al.*, 2011]. *Schweiger et al.* [2011] also emphasize the importance of model parameterizations such as an ITD or the ice strength and the difficulty in evaluating their impact. One way forward is to combine different datasets. For example, *Lindsay and Schweiger* [2015] used ice thickness observations from different sources to reduce the uncertainty in Arctic-wide trends; *Stroeve et al.* [2014] compared models of the Climate Model Intercomparison Project Phase 5 (CMIP5) with a similar collection of thickness data and showed that these models still cannot accurately reproduce statistics, regional distributions and trends of ice thickness; *Chevallier et al.* [2016] reported that observed concentrations are modelled accurately in global ocean reanalysis products, but that errors with respect to observed drift speeds remain and that there were large differences between the models in the regional ice thickness fields with no product standing out.

With the availability of data being a limiting factor, a common method to assess the impact of an ITD parameterization on sea ice models is to compare model configurations with and without this parameterization. *Bitz et al.* [2001] found in a coupled global climate model that including an ITD increases the mean ice thickness. This increase improved the fit to upward-looking sonar observations for mainly thick, ridged ice in the central Arctic, but deteriorated the fit in the peripheral seas. In addition, the interannual variability of both the sea ice export through Fram Strait and the ocean meridional overturning circulation increased with an ITD model. Feedback mechanisms were found to have a stronger effect on the sea ice in climate simulations with an ITD model [*Holland et al.*, 2006]. *Komuro and Suzuki* [2013] show the positive impact of this parameterization on the reproduction of realistic heat fluxes through the pack ice. *Maslowski and Lipscomb* [2003] compared two successive versions of a sea ice model and found that the later version improved the reproduction of sea ice observations significantly for which they stated the inclusion of an ITD parameterization into the model as the main reason. *Massonnet et al.* [2011] compared NEMO-LIM2 and NEMO-LIM3 model output to a much more exhaustive set of observations, but arrived at the same conclusions that the inclusion of an ITD parameterization into the model is one of the main reasons for a much improved model performance. All studies clearly show the positive impact of including an ITD model, but all evaluations are either limited by the lack of reliable observational data (again) or the simultaneous change of multiple model components confounds the conclusions.

Here we attempt a systematic investigation of the impact of an ITD parameterization on the reproduction of different large-scale observations of sea ice. We are supported by the ever increasing amount of available observational data. Our approach to systematic comparisons contains three steps: (1) We construct a cost function with error-weighted satellite data for sea ice concentration, thickness and drift as a robust measure of model performance; (2) We use this cost function to systematically tune different model configurations with and without an ITD model separately; that is, we explicitly do not use the same model parameters when using an ITD or a single-category model to avoid biases introduced by different parameterizations as much as possible. (3) We distinguish clearly between the effects of changing the ice thickness representation and the effects of changing the ice strength formulation.

The remainder of the paper is structured as follows: First we describe how we evaluate the different model configurations in section 2. This section contains an overview over the cost function, the optimization technique, the most important model equations, and the approach to tuning the different model configurations. The results of these comparisons are presented in section 3. The results are discussed in section 4 and the most important conclusions can be found in section 5.

2. Method

2.1. Cost Function

To evaluate our model results quantitatively we construct a cost function from satellite observations as a measure for model quality. We follow *Kauker et al.* [2015] and use four different datasets: (1) the reprocessed concentration dataset and error estimates from OSISAF [*EUMETSAT Ocean and Sea Ice Satellite Application Facility*, 2011] (1979 - 2009); (2) the ICESat-JPL thickness product [*Kwok and Cunningham*, 2008] with a local error estimated as in *Kauker et al.* [2015] yet with an upper limit of 1m for the uncertainty (March and October/November, 2003 - 2008); (3) the OSISAF sea ice drift [*Lavergne et al.*, 2010] (October to April, 2002 - 2006) and (4) the sea ice drift of *Kimura et al.* [2013] (May to July, 2003 - 2007). All of the drift data are derived from passive-microwave satellite data, with error estimates provided by *Sumata et al.* [2014, 2015].

The cost function F is defined as

$$F = \sum_{i=1}^{N} \frac{(y_i - x_i)^2}{N_d(y_i)\xi_i^2} \qquad (1)$$

where y_i is an observational data point with measurement uncertainty ξ_i, x_i the simulated value of the corresponding model variable, $N_d(y_i)$ the number of data points in each of the four datasets, and N the total number of observations. In equation (1) each data point y_i is weighted by $1/N_d$ in order to give equal weight to all four datasets. For instance, if the error for each data point $(x_i - y_i)$ was exactly equal to the measurement uncertainty ξ_i, the cost function for each dataset would be equal to one, summing up to a total value of $F = 4$. Note that the cost function is an average misfit of all included points, so that even for cost function values of less than four there can (and indeed do) exist regions where further improvement is still possible without overfitting.

2.2. Green's Function Approach

For a meaningful comparison of two model configurations, both configurations are tuned individually to minimize the differences between simulated and observed concentration, thickness and drift fields from 1979 to 2009. We use an semi-automatic optimization approach for a set of parameters with large impact on the ITD. The adjoint capabilities of the MITgcm [e.g. *Heimbach et al.*, 2010] cannot be used to optimally estimate the parameters, because our experiments span multiple decades. Instead we use Green's functions to linearize the problem and obtain a maximum likelihood estimate for a set of optimal parameters. A detailed mathematical background for the Green's function approach can be found in textbooks [e.g. *Menke*, 2012], while the short description below follows *Menemenlis et al.* [2005].

The relationship between the vector of observational data $\mathbf{y}$ and the model can be expressed as

$$\mathbf{y} = M(\boldsymbol{\nu}) + \boldsymbol{\varphi} \tag{2}$$

where the operator M combines the integration of the model and the sampling of the output at the specific locations. The model depends on a set of control parameters, for which $\boldsymbol{\nu}$ is a vector of perturbations around a reference $\boldsymbol{\nu_0}$. $\boldsymbol{\varphi}$ is the remaining error due to non-perfect parameter choices and systematic errors in the model. To get an optimal estimate of the control parameters $\boldsymbol{\nu_0} + \boldsymbol{\nu}$, a cost function

$$F = \boldsymbol{\varphi}^T \mathbf{R}^{-1} \boldsymbol{\varphi} \tag{3}$$

is minimized that measures a least-squares error weighted by a symmetric matrix $\mathbf{R}^{-1}$. For the special cost function (1) in section 2.1, the error is the model-data misfit $\varphi_i = y_i - x_i$ and $\mathbf{R}^{-1}$ is diagonal with elements $R_{ii}^{-1} = (\, N_d(y_i)\xi_i^2\,)^{-1}$. Equation (3) is minimized after linearizing operator M with a matrix $\mathbf{M}$. $\mathbf{M}$ is constructed by writing the Green's function for each of the control parameters into a new column. This first order approximation allows to write equation (2) as

$$\Delta\mathbf{y} = \mathbf{y} - M(\mathbf{0}) \;=\; \mathbf{M}\boldsymbol{\nu} + \boldsymbol{\varphi} \tag{4}$$

with the model data misfit $\Delta\mathbf{y}$. In this notation, $M(\mathbf{0})$ is the sampled output of a model integration with the reference set of control parameters $\boldsymbol{\nu_0}$, that is, the vector of perturbations is $\mathbf{0}$. Differentiating (3) with respect to the control vector $\boldsymbol{\nu}$ and equating the resulting gradient to zero, we obtain

$$\frac{\partial F(\boldsymbol{\nu}_{\text{opt}})}{\partial \boldsymbol{\nu}} = -\mathbf{M}^T\mathbf{R}^{-1} 2\left(\Delta\mathbf{y} - \mathbf{M}\boldsymbol{\nu}_{\text{opt}}\right) = 0. \tag{5}$$

Solving for the perturbation

$$\boldsymbol{\nu}_{\text{opt}} = \left(\mathbf{M}^T\mathbf{R}^{-1}\mathbf{M}\right)^{-1}\mathbf{M}^T\mathbf{R}^{-1}\Delta\mathbf{y} \tag{6}$$

gives a set of optimal control parameters $\boldsymbol{\nu_0} + \boldsymbol{\nu}_{\text{opt}}$. As a criterion for a successful optimization, the linearization error by this approach should be much smaller than the vector

$\boldsymbol{\xi}$ consisting of the measurement uncertainties ξ_i

$$\|M(\boldsymbol{\nu}_{\text{opt}}) - \mathbf{M}\boldsymbol{\nu}_{\text{opt}}\| \ll \|\boldsymbol{\xi}\|. \tag{7}$$

Because each of the Green's functions is calculated by one sensitivity experiment, the total computational effort necessary to construct $\mathbf{M}$ limits the number of control parameters.

2.3. Model Equations

2.3.1. Momentum Equations and Thermodynamics

For the dynamic part of the model we assume a viscous-plastic rheology with an elliptical yield curve and a normal flow rule [*Hibler*, 1979]. The ice velocities are calculated from the momentum balance:

$$m\frac{\partial \mathbf{u}}{\partial t} = m f_C \mathbf{k} \times \mathbf{u} + \boldsymbol{\tau}_a + \boldsymbol{\tau}_w - m\hat{g}\boldsymbol{\Delta}_H + \nabla \cdot \boldsymbol{\sigma}, \tag{8}$$

where $m = \rho_i h$ is the ice mass per unit area, h is the ice thickness, ρ_i is the ice density, $\mathbf{u}$ is the sea ice velocity vector, f_C is the Coriolis parameter, $\mathbf{k}$ is a unit vector pointing vertically upward, $\boldsymbol{\Delta}_H$ is the sea surface tilt, $\hat{g}$ is the gravitational acceleration and $\boldsymbol{\sigma}$ is the internal ice stress. The surface stress $\boldsymbol{\tau}_a$ and the water drag $\boldsymbol{\tau}_w$ can be written as

$$\tau_a = \rho_a C_a |\mathbf{u}_a - \mathbf{u}| \mathbf{R}_a (\mathbf{u}_a - \mathbf{u}) \tag{9}$$

$$\tau_o = \rho_o C_o |\mathbf{u}_o - \mathbf{u}| \mathbf{R}_o (\mathbf{u}_o - \mathbf{u}) \tag{10}$$

where $\mathbf{u}_a, \mathbf{u}_o$ are the surface velocities, ρ_a, ρ_o are the reference densities, C_a, C_o are the drag coefficients, and $\mathbf{R}_a, \mathbf{R}_o$ are rotation matrices for atmosphere (subscript a) and ocean (subscript o) [*McPhee*, 1975]. Following *Zhang and Hibler* [1997], the momentum balance (8) neglects the advection of momentum. The resulting discretized equations are solved using a line successive relaxation method [*Zhang and Hibler*, 1997].

The stress tensor $\boldsymbol{\sigma}$ is related to the deformation rate tensor $\dot{\boldsymbol{\varepsilon}} = \frac{1}{2}\left[\nabla \mathbf{u} + (\nabla \mathbf{u})^T\right]$ by the constitutive relation

$$\boldsymbol{\sigma} = 2\eta\dot{\boldsymbol{\varepsilon}} + \left((\zeta - \eta)\dot{\varepsilon}_I - \frac{P_r}{2}\right)\mathbf{I} \tag{11}$$

where P_r is the replacement pressure, $\mathbf{I}$ is the Identity Matrix, η and ζ are the shear and bulk viscosities, and $\dot{\varepsilon}_I = \dot{\varepsilon}_{11} + \dot{\varepsilon}_{22}$ is the first strain rate invariant (i.e. divergence). The bulk viscosity $\zeta = P/(2\Delta_{\dot{\varepsilon}})$ and the shear viscosity $\eta = \zeta/e^2$ in turn can be calculated from the ice strength P, the axis ratio e of the elliptical yield curve, and the deformation measure $\Delta_{\dot{\varepsilon}} = \sqrt{\dot{\varepsilon}_I^2 + e^{-2}\dot{\varepsilon}_{II}^2}$, where $\dot{\varepsilon}_{II} = \sqrt{(\dot{\varepsilon}_{11} - \dot{\varepsilon}_{22})^2 + 4\dot{\varepsilon}_{12}^2}$ is the second strain rate invariant (or maximum shear at a point). The replacement pressure $P_r = 2\Delta_{\dot{\varepsilon}}\zeta$ is calculated after regularizing ζ with the smooth formulation of *Lemieux and Tremblay* [2009] to avoid spurious creep [*Hibler and Ip*, 1995].

The single-category model is based on the two continuity equations

$$\frac{\partial A}{\partial t} = -\nabla \cdot (\mathbf{u}A) + S_A \tag{12}$$

$$\frac{\partial H}{\partial t} = -\nabla \cdot (\mathbf{u}H) + S_H \tag{13}$$

for the prognostic variables ice concentration A and ice volume per grid cell area $H = Ah$. The variables change with time according to advection by the horizontal velocity $\mathbf{u}$ and the respective source terms S_A and S_H. The thermodynamic fluxes are calculated using a 0-layer model [*Semtner*, 1976]. Note that *Bitz et al.* [2001] analyzed the impact such simple thermodynamics have on an ITD model compared to more complex thermodynamics. They found that ice concentration is almost indistinguishable from the one simulated with more complex thermodynamics but there are non-negligible changes in ice thickness and growth rates, which should be kept in mind for the interpretation of the results presented below.

2.3.2. Ice Thickness Distribution

One main focus of our investigation is the subgrid-scale ice thickness distribution $g(h, \mathbf{x}, t)$ [*Thorndike et al.*, 1975], a probability density function for thickness h following the evolution equation

$$\frac{\partial g}{\partial t} = -\nabla \cdot (\mathbf{u}g) - \frac{\partial}{\partial h}(fg) + \Psi, \tag{14}$$

where f is the thermodynamic growth rate and Ψ a function describing the mechanical redistribution of sea ice during ridging or lead opening.

The mechanical redistribution function Ψ creates open water when the sea ice flow is divergent and ridges when the sea ice flow is convergent. The function Ψ depends on the total strain rate and the ratio between shear and divergent strain. In convergent motion, the ridging mode

$$\omega_r(h) = \frac{n(h) - a(h)}{N} \tag{15}$$

gives the effective change of ice volume for thickness between h and $h + \mathrm{d}h$ as the normalized difference between the ice $n(h)$ generated by ridging and the ice $a(h)$ participating in ridging. Following *Lipscomb et al.* [2007], the participation function is $a(h) = b(h)g(h)$, and the relative amount of ice of thickness h is weighted by an exponential function

$$b(h) = b_0 \exp[-G(h)/a^*], \tag{16}$$

where $G(h) = \int_0^h g(h)\mathrm{d}h$ is the cumulative thickness distribution function, b_0 is a normalization factor, and a^* determines the relative amount of thicker and thinner ice that take part in ridging. The ice generated by ridging (from an original thickness h_1 to a new ice thickness h) is calculated as

$$n(h) = \int_0^\infty a(h_1)\gamma(h_1, h)\mathrm{d}h_1, \tag{17}$$

where the density function $\gamma(h_1, h)$ can be written as:

$$\gamma(h_1, h) = \begin{cases} \frac{1}{k\lambda} \exp\left[\frac{-(h-h_{\min})}{\lambda}\right] & h \geq h_{\min} \\ 0 & h < h_{\min}. \end{cases} \tag{18}$$

In this parameterization, the normalization factor $k = \frac{h_{\min}+\lambda}{h_1}$, the e-folding scale $\lambda = \mu h_1^{1/2}$ and the minimum ridge thickness $h_{\min} = \min(2h_1, h_1 + h_{\text{raft}})$ all depend on the original thickness h_1. The maximal ice thickness allowed to raft is constant $h_{\text{raft}} = 1$m and μ is a tunable parameter.

In the numerical implementation these equations are discretized into a set of thickness categories using the delta function scheme proposed by *Bitz et al.* [2001]. A smoother linear remapping scheme [*Lipscomb*, 2001] is available but not used. Its effect will be discussed in section 4.1. For each thickness category in an ITD configuration, the volume conservation law equation (13) is evaluated as in the single-category model, but with the net surface ice-atmosphere heat flux calculated from the values for ice and snow thickness in the current category. There are no conceptual differences in the thermodynamics between the single-category and ITD configurations. The only difference is that in the ITD configuration, new ice of thickness H_0 is created only in the thinnest category; all other categories are limited to basal growth. The conservation of ice area (12) is replaced by the discretized evolution equation for the ITD (14). The thickness category limits of the discretization in space are given in Table 1. The total ice concentration and volume can then be calculated by summing up the values for each category.

In the single-category model ridge formation is treated implicitly by limiting the ice concentration to a maximum of one [*Hibler*, 1979]. In this simple case ($A = 1$), the concentration can no longer increase and convergence leads then to an increase in ice thickness (i.e. a “ridge”).

2.3.3. Ice Strength Parameterizations

Rothrock [1975] derived a parameterization for the ice strength P

$$P = C_f C_p \int_0^\infty h^2 \omega_r(h) \mathrm{d}h \tag{19}$$

from considerations of the amount of potential energy gained and frictional energy dissipated during ridging. The physical constant $C_p = \rho_i(\rho_w - \rho_i)\hat{g}/(2\rho_w)$ is a combination of the gravitational acceleration $\hat{g}$ and the densities ρ_i, ρ_w of ice and water, and C_f is a scaling factor relating the work against gravity to the work against friction during ridging.

Hibler [1979] proposed a simpler ice strength parameterization for a single-category model that is still widely used today. In this model the ice strength P is parameterized as

$$P = P^* A\, h\, e^{-C^*(1-A)} \tag{20}$$

where P depends only on average ice concentration and thickness per grid cell, the compressive ice strength parameter P^* and the ice concentration parameter C^*. In the following we will refer to the ice strength parameterization of *Hibler* [1979] as H79 and that of *Rothrock* [1975] as R75.

Note that the parameterization R75 is a function of the ITD in each grid cell, while H79 is applicable both for ITD and single-category models. In contrast to H79, which builds on the plausible assumption that thick and compact ice has more strength than thin and loosely drifting ice, the R75 parameterization clearly contains more physical assumptions about energy conservation. For that reason R75 is often considered to be more physically realistic than H79.

2.4. Optimization Approach

2.4.1. Optimized Parameters

We define three groups of control parameters for our optimization that we think are most important for adjusting the modeled sea ice to observations. Group 1 contains parameters that are not directly related to the choice of ITD parameterizations: the albedo of cold and melting snow and ice, the air and water drag coefficients, the aspect ratio e of the elliptical yield curve, and the thickness of newly formed ice H_0. Group 2 contains parameters only relevant to the H79 ice strength formulation: the ice compressive strength parameter P^* and the ice concentration constant C^*. Finally group 3 contains parameters of the R75 strength formulation: the ice strength parameter C_f, and the ice redistribution coefficients μ and a^*.

2.4.2. Optimization Runs

For our comparisons we have three goals in mind: (1) evaluate the differences of model configurations with and without an ITD with respect to reproducing observed sea ice fields; (2) account for the influence of the number of ice thickness categories; (3) account for the influence of the ice strength parameterization. The quality of each model configuration is measured by means of a cost function. For an unbiased comparison of model quality, we first tune each model configuration in order to minimize the total cost function F.

We use the MIT general circulation model (MITgcm), in a coupled ocean / sea-ice configuration, forced with prescribed atmospheric reanalysis data. In this configuration, which is a coarser version of *Nguyen et al.* [2011], we implemented the ITD model in the MITgcm sea ice model [*Losch et al.*, 2010]. The model region is the Arctic face of a global cubed sphere configuration with an average resolution of 36 km. Similar sea ice models are currently being used in configurations with horizontal resolutions between 5 km for regional simulations [*Dupont et al.*, 2015] and around 50 km for global reanalysis [*Chevallier et al.*, 2016]. Our model is therefore representative of a broad group of medium resolution models. All model runs start from a 5-year spinup with periodic forcing of the year 1979. The model is then run from 1979 to 2009.

The initial choice of model parameters follows *Nguyen et al.* [2011], but we use a more recent atmospheric forcing data set following the recommendations of *Lindsay et al.* [2014]: The NCEP Climate System Forecast Reanalysis [NCEP-CSFR *Saha et al.*, 2010] produced the best results for our configuration in a comparison of different reanalysis products (i.e. the smallest model-data misfit prior to the formal optimization, not shown).

Starting from the tuned set of parameters of *Nguyen et al.* [2011], we adjust the parameters of group 1 with one optimization step to account for the differences in forcing, grid resolution and other model details. This setup without ITD parameterization is referred to as the “Baseline” hereafter. Next we tune a case with an ITD using five ice thickness categories, a number recommended by *Bitz et al.* [2001]. In order to determine the parameters to be adjusted when switching to an ITD, we perform three different optimizations with the non ITD specific parameters of group 1 (“ITD5-g1”), the ITD and R75 specific parameters of group 3 (“ITD5-g3”) or both sets together (“ITD5-g13”). Table 2 lists which parameters are modified in which experiment. The best result (minimum cost function F) is obtained when only tuning the ITD specific parameters of group 3 (Table 3). Therefore we continued from Baseline by tuning parameters of group 3 for two different numbers of ice thickness categories (5 and 20) with the R75 ice strength parameterization to arrive at the configurations ”ITD5R“ and ”ITD20R“.

Tuning the strength-specific parameters of group 2 yields the configuration noITD with a single-category thickness representation. In order for those optimizations to satisfy criterion (7), we require the linearization error to be smaller than 10% of the observation uncertainty on average. This requirement was satisfied in one step for noITD and two steps for each of ITD5R and ITD20R. This optimization approach decreases the cost function values of the ITD configurations by $25\% - 30\%$ (Table 3).

To assess the role of the strength parameterization in the context of an ITD model, we evaluated two additional model runs with an ITD and the simpler H79 ice strength parameterization: "ITD5H" and "ITD20H". For those runs we assume that the parameters, which have already been tuned using our cost function, give sufficiently good results in this new combination. Therefore we forego further optimization for the runs ITD5H and ITD20H and instead use the parameters from the respective R75 runs with the values P^* and C^* from noITD.

This approach implies that the thickness of newly formed ice is $H_0 = 0.5649$, the value resulting from the optimization of the Baseline configuration, in all ITD configurations. Arguably, this high value may prevent the ITD model from representing the behavior of thin ice realistically, especially since the thinnest category for ITD20 contains only ice thinner than 16 cm. To investigate the effect of this artifact on our analysis, we additionally optimize only H_0 for the two configurations ITD5R and ITD20R. We find that it is possible to further decrease the model-data misfit by tuning H_0 as shown in Table 3 for runs "ITD5R-H0" and "ITD20R-H0", but that our qualitative results are not affected. Tuning of H_0 also does not reduce the value of H_0 to be within the limits of the thinnest category for ITD20R (see Table 2). We thus conclude that it is not necessary to contain newly formed ice in the thinnest thickness category in order to minimize model-data differences. An overview of the different optimized runs is given in Table 4.

3. Results

Based on the cost function, both combinations of ITD and H79 give best results and even the configuration noITD has a smaller cost function value than the two configurations with ITD and R75. This result is described in more detail in section 3.1. We then investigate separately the influence of the ITD (section 3.2) and the strength parameterization (section 3.3) on the quality and characteristics of the model results in order to explain why the configurations with R75 have difficulties fitting the data. Especially for the ice strength parameterization, we find a strong dependence on the thickness resolution in the ITD. For this reason, we account for the different number of thickness categories throughout this section.

The simulated sea ice climate in our experiments is very close to the one described by *Nguyen et al.* [2011]. Due to our more specific tuning, we can even improve the fit to sea ice observations compared to their already very good model state, but still suffer from biases in thickness and concentration, that are common to many comparable models [*Chevallier et al.*, 2016]. We therefore assume that our model provides a good representation of Arctic sea ice and we focus our analysis on the differences in the fit to observations, as expressed by our cost function, that are caused by changes in the model setup.

3.1. Cost function

The total error calculated from the cost function F is slightly larger for both ITD5R and ITD20R when compared to noITD and significantly larger than both model configurations ITD5H and ITD20H. An overview of the cost function values can be found in Table 3.

To investigate the individual strengths and weaknesses of the different model configurations in more detail, we split up the total cost function values into four contributions for each of the individual datasets (Table 3). The difference between the four different ITD configurations (ITD[5,20][R,H]) and noITD are shown in Figure 1. The ITD configurations using R75 improve the fit to some datasets, but this reduction in cost function is outweighed by increases in differences in others. For instance, ITD5R has a clearly better fit to concentration data than noITD and a slightly better fit to thickness, but the fit to the drift data is much worse than in noITD. ITD20R, on the other hand, has in total a comparable and in winter even a slightly better fit to the drift data than noITD, but the fit to thickness and concentration is much worse compared to ITD5R. Part of this behavior can also be observed for ITD5H and ITD20H: In this case the fit to thickness and drift is similar, but the fit to concentration is much better for ITD5H than for ITD20H. These observations are a first hint of the strong influence of the number of thickness categories on the simulated sea ice concentration for a general ITD model, but also on all other sea ice characteristics for the R75 strength parameterization.

3.2. ITD

We isolate and assess the effect of the ITD model by first comparing the configuration noITD with ITD5H and ITD20H, all of which use the same strength parameterization H79.

The more complex ITD model reduces the misfit for ice concentration especially in the marginal ice zone for the entire year, see Figure 2 for summer results; winter results are not shown. All model configurations generally overestimate the concentration especially in the North Atlantic, where the ice edge extends too far south and south east. While this overestimation is found in many medium resolution models [*Chevallier et al.*, 2016], the ITD configurations largely reduce this misfit when compared to noITD. In contrast, the summer ice concentration in the central Arctic and in the straits of the Canadian Arctic Archipelago is higher with an ITD model (Figure 2). This is because most ice in the ITD model is in the thicker ice categories and thicker ice takes longer to melt. In the noITD model, sea ice melt leads to sea ice concentration changes even for thicker ice because a linear ice thickness distribution between 0 and $2h$ is assumed so that there is always thin ice available for fast melting.

The ice thickness generally increases with number of ice thickness categories, with much stronger tendencies in the straits of the Canadian Arctic Archipelago. The difference in ice thickness between ITD5H - noITD is $0.11 \pm 0.20\,\mathrm{m}$ (mean and standard deviation) for ice thinner than 4 m in ITD5H, and the comparable difference between ITD20H - noITD is $0.17 \pm 0.25\,\mathrm{m}$. These differences grow to $1.14 \pm 1.67\,\mathrm{m}$ for ITD5H and $1.45 \pm 1.49\,\mathrm{m}$ for ITD20H, if only ice thicker than 4 m in the ITD run is taken into account. Ice of this thickness is found mainly in the straits of the Canadian Arctic Archipelago and north of Greenland.

We now explicitly compare the ITD5 and ITD20 configurations for both strength parameterizations R75 and H79 in order to investigate the impact of the number of thickness categories. For ITD20 we observe generally a larger total ice volume compared to ITD5:

First, if there is ice in an ITD5 configuration with a concentration of less than one, the concentration is in almost all cases higher in the corresponding ITD20 run. Second, the higher thickness observed for an ITD model compared to noITD is further increased, with the differences between ITD20 and ITD5 (Figure 3) showing a similar pattern as the differences between an ITD5 configuration and noITD (not shown).

The differences in ice drift are less clear. We find mostly higher drift speeds in the configurations ITD20R than in ITD5R, while we find the exact opposite for ITD20H and ITD5H. This ambiguous result can be explained by the effect of ice thickness resolution on the ice strength parameterization (see subsection 3.3, below).

3.3. Ice Strength

In this section, the effects of the different strength parameterizations on an ITD model are compared in greater detail. In this context, the role of the number of thickness categories is emphasized.

We find that the non-linearity in the R75 parameterization leads to higher fluctuations in the ice strength on the near-grid scale. For both ITD5 and ITD20, the most prominent difference between the strength formulations is found in the ice thickness of very thick ice north of Greenland and the Canadian Archipelago. Ice exceeding four meters in thickness, which mainly exists in those regions, is on average thicker by more than seventy centimeters in the R75 runs when compared to H79; but ice thinner than two meters, especially common in the peripheral regions of the Arctic, is slightly thinner on average with R75 when compared to H79 (Figure 4). As a possible explanation for these observations, we see generally larger ice strength gradients with R75 than with H79, with the most prominent differences north of Greenland and Ellesmere Island (results not shown). The calculation of the ice strength following R75 depends non-linearly on the local distribution of ice into different thickness categories, so that to some degree higher small-scale fluctuations are expected. But the magnitude of those strength gradients can lead to stronger gradients in the velocity fields, especially for otherwise immobile ice. Due to this process we find in the runs using R75 higher convergence rates for ice thicker than 3 m (Figure 5). This increased ridging especially in regions of already thick ice dynamically creates peak ice thicknesses much higher than observed.

The differences in concentration and drift between R75 and H79 are less clear for all ITD configurations. The differences in sea ice concentration for ITD5 and ITD20 for a climatological August are plotted in Figure 6; the patterns are very similar throughout the year. The ice in the marginal ice zone between Siberia and Svalbard, in winter and spring even down to Iceland, is less compact for R75 than for H79. At the same time, the ice concentration is larger for R75 in the other marginal seas, most notably in the Beaufort and Chukchi Seas and in the Baffin Bay. In the central Arctic, the differences in concentration depend on the number of thickness categories: in the ITD5 configurations, the ice is more compact for R75 than H79; but in the ITD20 configurations, the ice in summer is slightly less compact for R75 compared to H79. The ice drift is slower for R75 in large parts of the central and western Arctic and faster in the outflow of the transpolar drift and in Fram Strait (not shown). In the remaining Arctic regions we find a similar ambiguity as in the concentration fields: For R75, the ice tends to be slightly slower in the ITD5 configurations and slightly faster in the ITD20 configurations when compared to H79. Those changes can be traced back to similar patterns in the ice strength with the ice being weaker for R75 where it is faster and vice versa (not shown).

We explain those differences by the effects of two different mechanisms. On the one hand, the mean ice state with R75 is characterized over large parts of the central and western Arctic by larger thicknesses and often also slightly higher concentrations. Physically, those changes in the mean ice state generally lead to higher ice strength and thereby slower drift. On the other hand, the ice strength is a non-linear function of thickness distribution for R75, which makes the differences to the linear H79 formulation not uniform. To illustrate this we compare the strength values for both R75 and H79 computed from the ice states of model simulations using R75. For ice with a compressive strength (R75) higher than $40,000\,\mathrm{Nm}^{-2}$, the strength values calculated by R75 are higher than those for H79, and the differences grow linearly with the ice strength over a large range (Figure 7). In contrast, in the range below $30,000\,\mathrm{Nm}^{-2}$, the ice strength values calculated by R75 are lower than those for H79.

Finally, the R75 ice strength depends more strongly on the actual distribution of ice thicknesses than on the averaged characteristics of the sea ice. Figure 8 shows the difference in ice strength together with the difference in ice thickness between ITD5 and ITD20 simulations for both strength parameterizations. The ice thickness is mainly larger for the ITD20 model for both H79 and R75. As expected following the simple relationship (20) and the physical understanding that thicker ice is more difficult to deform, H79 calculates higher ice strength for the thicker ice in ITD20 over most thickness bins. The impact of the ice thickness on the ice strength reduces for ice thicker than three meters, most likely because of the increasing effect of the replacement pressure method [*Hibler and Ip*, 1995], which tends to reduce the ice strength of thick, immobile pack ice. In contrast, while for R75 the mean thickness is also mostly higher in the ITD20 configuration than in ITD5, the average ice strength is lower. So for this ice strength formulation, finely resolving the thin ice categories (and thereby weakening the ice pack) has a larger impact on the ice strength than the physical property that thicker ice should be more difficult to deform.

4. Discussion

The H79 ice strength formulation can be justly criticized because it is not derived from first principles. Therefore, the option of using the physically motivated R75 formulation is often thought of as a great advantage of an ITD model. In contrast to that notion, our results suggest that simulating realistic drift fields with medium-resolution sea ice models with R75 strength is difficult. In particular, in our simulations the model performance did not improve over a sufficiently tuned single-category set-up after including an ITD parameterization together with the commonly used R75 strength parameterization. Somewhat counter-intuitively, the model performance was better for fewer thickness classes and the model especially improved when the ITD was combined with the H79 strength formulation.

4.1. ITD

Our model overestimates the concentration along the ice edge almost everywhere in the North Atlantic and most of the time. In both ITD5 runs this overestimation is greatly reduced. *Bitz et al.* [2001] described a similar effect and explained it by faster melting of thin categories in the ITD, which leads to more open water, that is, lower ice concentration, especially during the summer season. Somewhat in contrast, we find also higher summer ice concentrations for the ITD configurations, mostly in the central ice pack. We explain this also by the same effect of thin ice melting. The single-category approach of *Hibler*

[1979] assumes a uniform distribution of thickness between 0 and $2h$ for the creation of open water, so that there can be more thin ice available in this configuration than in the ITD models, which may not have any ice in the thinnest category.

In addition, the effect of an ITD model on the ice edge depends strongly on the number of categories. Resolving the ice thickness distribution better (ITD20 vs. ITD5 configurations) leads to higher ice concentrations in the marginal ice zone with the consequence of a larger ice edge position error than in the noITD model. We find that the increase in total ice volume and the associated ice export with more thickness classes is too strong to be balanced by the increased melting in the marginal ice zone that one would expect when the thinner categories are better resolved.

The mean ice thickness increases with the number of thickness classes (noITD < ITD5 < ITD20) [see also *Holland et al.*, 2006; *Komuro et al.*, 2012]. This result is consistent with the physical reasoning that a better resolution of thin ice in the pack allows for more ice growth, because heat fluxes and deformation (ridging) increase. In contrast, *Massonnet et al.* [2011] found in a comparison between model versions a decrease in ice thickness, which they attributed to the use of an ITD model. We argue, that their analysis may have been confounded because in comparing different model versions they changed multiple model components and parameters, including a lower value for the thickness of new ice H_0 in the model version with the ITD, which also changes ice thickness and concentration fields.

We did not fully address the question of (numerical) convergence of the ITD model with the number of thickness classes. A fine resolution of the thin ice range was found to be necessary to reproduce observed heat fluxes [*Bitz et al.*, 2001] and a better resolution of the upper thickness range was required to reproduce total ice volume [*Hunke*, 2014]. Based on our experiments with 5, the minimum number recommended by *Bitz et al.* [2001], and 20 classes, which were chosen to have a simulation with a nearly converged ITD model [*Lipscomb*, 2001], we find that the better resolved solution does not lead to the best model-data fit. More thickness classes increase the ice volume and eventually lead to an overestimation of thickness, apparently introducing a stronger bias in the solution than the effects of a coarse thickness resolution. It is unclear in how far these effects can be moderated by more realistic thermodynamics, as the thermodynamics can have a strong impact on ice thickness [*Bitz et al.*, 2001; *Losch et al.*, 2010].

The delta function scheme [*Bitz et al.*, 2001], which we use in our simulations, was criticized to be prone to produce numerical discontinuities in the ITD and to leave many thickness categories empty, thereby artificially reducing the thickness resolution [*Lipscomb*, 2001]. A linear remapping scheme was implemented to overcome these issues [*Lipscomb*, 2001]. We observe the same improvements in test simulations with the linear remapping scheme (smoother thickness distributions with fewer gaps, not shown), but also on average slightly thicker ice and higher ice concentration. The main results of our study, however, remain intact: the quality of the model output, measured by the cost function, is higher for ITD configurations with H79 than for noITD, which in turn is better than the combinations of ITD and R75; and notably we observe the same dependency of the ice strength on the number of thickness categories (not shown).

4.2. Ice Strength

Bitz et al. [2001] found that for R75 the ice is weaker if a given thickness distribution is better resolved. This is probably so because the strength of the ice pack is determined mostly by the amount of thin ice and if the thin end of the thickness distribution is better resolved, thinner ice can lead to smaller ice strength. H79 misses this sensitivity to thin ice because of linearity. We show that for R75 this effect can be strong enough in a realistic model set-up to outweigh the opposing effect of thicker ice resulting from more thickness categories (Figure 8). Although this behavior may be physical and could be seen as an advantage of R75 over H79, it reduces the ability to reproduce large-scale satellite observations in our experiments.

The differences in modeled ice drift patterns in our simulations are mostly caused by the different ice strength formulations, because other drivers such as the wind forcing were the same for all experiments. Because the number of thickness categories has such a strong impact on the solutions with R75, we cannot distinguish a clear change of drift patterns due to an ITD that would be independent of the choice of strength parameterization. In a comparison of different ocean-sea ice reanalysis products to satellite observations of ice drift — unfortunately they used a different observational data set, which makes a direct comparison of their results to ours difficult — *Chevallier et al.* [2016] identified the choice of atmospheric forcing and differences in drag coefficients as the most important model parameters and confirmed the strong role of the wind stress in determining the drift patterns of sea ice [*Hunke et al.*, 2011]. Our results indicate that when those leading-order effects are held constant, changing the formulation of ice strength is a powerful way of affecting the model-data misfit for sea ice drift.

Holland et al. [2006] attributed the increased ice thickness with an ITD model to the larger ice growth rates generally produced by an ITD. We can now distinguish the effects of the strength parameterization from the choice of thickness representation in the model to show that while an ITD leads to a general increase in the overall thickness, the choice of R75 is mainly responsible for excessively large maximal thicknesses north of Greenland and Ellesmere Island. These are caused by the strong small-scale gradients in the ice strength for R75 that allow higher deformation rates in very thick ice, so that already thick ice can be ridged further, eventually leading to much higher maximal thickness values than observed.

Although the derivation of R75 is arguably more physical than that of H79, it leads to a poorer model-data misfit. In the following we speculate about the reasons for this counter-intuitive result: *Rothrock* [1975] already mentioned two issues with known energy sinks in his derivation of the work necessary for ridge formation: (1) fracturing of ice was neglected following an argument of *Parmerter and Coon* [1973] and (2) frictional loss in shearing was neglected and assumed to be at most of the order of frictional losses in compression based on the notion of a Coulomb friction model. To estimate the work against friction in compression, *Rothrock* [1975] made strong assumptions about complicated processes of ice interaction without having enough data available to constrain them. He arrived at approximately similar contributions by gravitational and frictional work. This lead to a scaling factor $C_f = 2$, but later *Flato and Hibler* [1995] estimated this factor to be $C_f = 17$ based on a model comparison to observed buoy drift patterns. This large difference in C_f between estimates by theory and numerical model comparisons together with a re-evaluation of energy dissipation in shear [*Pritchard*, 1981] suggest to us that important physical effects are not properly included in the approach of R75.

Fundamental questions about the form of a new ice strength parameterization are unclear. For example, *Hopkins* [1998] found in model simulations of ridging processes that pressure ridge formation leads to a scaling of the ice strength proportional to $h^{3/2}$. *Hibler* [1980] also supports a scaling with $h^{3/2}$ by physical reasoning, but in the absence of sufficient observational data his theory is based in important parts on physical intuition. Note, however, that *Hopkins* [1998] considers only ice breaking in flexure, not in crushing. The load that ice can withstand before it is crushed grows linear in h [*Rothrock*, 1975]. Further, ice strength scaling with h^2 was found in numerical simulation of ridge formation with a different experimental set-up [*Hopkins et al.*, 1991]. The R75 ice strength scales with $h^{3/2}$, while the ice strength after H79 is linear in the mean thickness h [*Lipscomb et al.*, 2007], but neither appear to cover all observational evidence. We emphasize that there still exists great uncertainty in the exact nature of such a scaling. Our results indicate that the linear relationship [*Hibler*, 1979] might be better suited to represent Arctic-wide averages.

4.3. Qualitative Assessment of Our Results

Measuring the quality of our model results with the cost function (1) allows us to assess the overall performance of a given configuration in a detailed and quantifiable way. To this end, we evaluate the reproduction of large-scale sea ice features, such as sea ice extent, thickness and drift — as opposed to the details of the ocean state. Three of the four data products (thickness and both drift products) are limited to certain seasons in a few years, and two of them (thickness and drift from *Kimura et al.* [2013]) are also limited to the central Arctic. Still the combination of the four products allows a year-round coverage of the whole Arctic in those years. In our analysis, we implicitly assume that large errors in one sea ice property (e.g. thickness) would affect other sea ice properties (e.g. drift and concentration) in a detectable manner. Additionally, the availability of the concentration data for the entire thirty-year simulation period provides some measure against overfitting the model to the short period 2002 - 2008 covered by the other satellite products.

Are the results presented in section 3 sensitive to the exact choice of observations included in the cost function? We tested this by evaluating the cost function for any combination of three (out of four) sets of observations and found that the main conclusion of the paper is robust to the exact choice of observations. In all cases, the ITD configurations together with the strength parameterization H79 lead to a better fit to the observations than the single-category configuration noITD with the strength parameterization H79. The noITD case in turn leads to a better fit than the ITD with the ice strength parameterization R75 (Table 3).

Our modeling approach is based on a simple single-category ice model (in fact, it is a two-category model: ice and no-ice [*Hibler*, 1979]) without internal heat capacity (linear internal temperature profile) and without considering a brine parameterization [*Bitz and Lipscomb*, 1999]. Both of these omissions will lead to a larger seasonal amplitude in ice thickness and to the absence of a lag between the net surface heat fluxes and the seasonal cycle of ice thickness. When we minimize the cost function (1), the biases in ice thickness will be compensated by adjustments in the optimal choice of surface albedo for sea ice and snow. While it is true that we are compensating for a winter bias in ice growth (induced by the lack of thermal inertia) by including another bias in summer melt (via the albedo), the fact that we are mainly interested in the ice strength parameterization — something that is important only during one season (mid to late winter) when the ice interactions are significant [*Steele et al.*, 1997; *Richter-Menge*, 1997] — suggests that our conclusions are not sensitive to the presence or absence of sea ice thermal inertia. Moreover, the absence of a lag between surface atmospheric forcing and sea ice thickness will only be important for a few weeks near the onset of the melt season (the delayed ice growth in fall occurs at a time when the ice interactions are small, [*Richter-Menge*, 1997]); this will therefore result in second order changes in the cost function over the full winter season. For these reasons, we believe that the simpler treatment of thermodynamic will not impact the main conclusions.

The choice of forcing data generally has a large impact on model results [*Lindsay et al.*, 2014]. Prior to optimization, we chose the best forcing data set based on our cost function. A different forcing data set may change the magnitude of ice thickness or the regional distribution of ice and it will guide the optimization to a different set of optimized parameter values, but the internal mechanics of the model that are responsible for the differences between the parameterizations are not affected.

5. Conclusions

A rigorous model-data comparison for an ITD model and two different strength parameterizations leads us to the following conclusions: Sea ice models with an ITD parameterization can outperform single-category models in reproducing observed concentration, thickness, and drift fields. Somewhat unexpectedly, the best fit to observations is achieved with an ITD model following *Thorndike et al.* [1975] combined with a simple ice strength parameterization [*Hibler*, 1979]. The more sophisticated ice strength parameterization by *Rothrock* [1975] leads to the poorest agreement to observations, even compared to the single-category model: Problems associated with this parameterization over-compensate the positive effect of an ITD model on the overall model.

It is not obvious why the Arctic-wide behavior of sea ice is reproduced with the least accuracy for the ice strength parameterization after *Rothrock* [1975] in our simulations. We found the modeled physics to produce implausibly large peak ice thicknesses, probably due to very high deformation of already thick ice and also a very strong dependence of the modeled ice strength on the number of thickness categories. This points to potential issues in both the physical assumptions in the formulation and the numerical discretization procedure. A short term improvement may be achieved by using the ITD parameterization together with the H79 strength formulation for medium resolution models. But because of the lack of physical justification for this parameterization, this short-term solution may turn out to be insufficient for sea ice simulations in climate change scenarios.

The increasing availability of satellite data make possible detailed, quantitative analyses of model parameterizations. These can be further enhanced by additional data sources such as EM-Bird thickness measurements [*Haas et al.*, 2009] or ice age [*Hunke*, 2014]. We argue that in order to realistically reproduce Arctic sea ice it is necessary to re-evaluate the ice strength formulation as a major link between ice volume and ice drift.

Acknowledgments. We would like to thank Frank Kauker for sharing the preprocessed data sets and for help with setting up the cost function. And we would like to thank Jean-François Lemieux and an anonymous reviewer for their helpful comments on this paper. This project was supported by the Deutsche Forschungsgemeinschaft (DFG) through the International Research Training Group "Processes and impacts of climate change in the North Atlantic Ocean and the Canadian Arctic" (IRTG 1904 ArcTrain). The source code for the model used in this study, the MITgcm, is freely available at

http://mitgcm.org. The model output evaluated in this paper is archived in PANGAEA and available at https://doi.org/10.1594/PANGAEA.865445

References

Bitz, C. M., and W. H. Lipscomb (1999), An energy-conserving thermodynamic model of sea ice, *Journal of Geophysical Research*, *104*(C7), 15,669, doi:10.1029/1999JC900100.

Bitz, C. M., M. M. Holland, A. J. Weaver, and M. Eby (2001), Simulating the ice-thickness distribution in a coupled climate model, *Journal of Geophysical Research*, *106*, 2441, doi:10.1029/1999JC000113.

Chevallier, M., G. C. Smith, F. Dupont, J.-F. Lemieux, G. Forget, Y. Fujii, F. Hernandez, R. Msadek, K. A. Peterson, A. Storto, T. Toyoda, M. Valdivieso, G. Vernieres, H. Zuo, M. Balmaseda, Y.-S. Chang, N. Ferry, G. Garric, K. Haines, S. Keeley, R. M. Kovach, T. Kuragano, S. Masina, Y. Tang, H. Tsujino, and X. Wang (2016), Intercomparison of the Arctic sea ice cover in global oceansea ice reanalyses from the ORA-IP project, *Climate Dynamics*, *SI*, doi:10.1007/s00382-016-2985-y.

Dupont, F., S. Higginson, R. Bourdallé-Badie, Y. Lu, F. Roy, G. C. Smith, J.-F. Lemieux, G. Garric, and F. Davidson (2015), A high-resolution ocean and sea-ice modelling system for the Arctic and North Atlantic oceans, *Geoscientific Model Development*, *8*(5), 1577–1594, doi:10.5194/gmd-8-1577-2015.

EUMETSAT Ocean and Sea Ice Satellite Application Facility (2011), Global sea ice concentration reprocessing dataset 1978-2009 (v1.1), [Online]. Norwegian and Danish Meteorological Institutes. Available from http://osisaf.met.no.

Flato, G. M., and W. D. Hibler (1995), Ridging and strength in modeling the thickness distribution of Arctic sea ice, *Journal of Geophysical Research*, *100*(C9), 18,611–18,626.

Godlovitch, D., R. Illner, and A. Monahan (2011), Smoluchowski Coagulation Models Of Sea Ice Thickness Distribution Dynamics, *Journal of Geophysical Research*, pp. 1–40.

Haas, C., J. Lobach, S. Hendricks, L. Rabenstein, and A. Pfaffling (2009), Helicopter-borne measurements of sea ice thickness, using a small and lightweight, digital EM system, *Journal of Applied Geophysics*, *67*(3), 234–241, doi:10.1016/j.jappgeo.2008.05.005.

Heimbach, P., D. Menemenlis, M. Losch, J.-M. Campin, and C. Hill (2010), On the formulation of sea-ice models. Part 2: Lessons from multi-year adjoint sea-ice export sensitivities through the Canadian Arctic Archipelago, *Ocean Modelling*, *33*(1), 145–158.

Herzfeld, U. C., E. C. Hunke, B. W. McDonald, and B. F. Wallin (2015), Sea ice deformation in Fram Strait Comparison of CICE simulations with analysis and classification of airborne remote-sensing data, *Cold Regions Science and Technology*, *117*, 19–33, doi:10.1016/j.coldregions.2015.05.001.

Hibler, W. D. (1979), A Dynamic Thermodynamic Sea Ice Model, *Journal of Physical Oceanography*, *9*(4), 815–846, doi:10.1175/1520-0485(1979)009¡0815:ADTSIM¿2.0.CO;2.

Hibler, W. D. (1980), Modeling a Variable Thickness Sea Ice Cover, *Monthly Weather Review*, *108*(12), 1943–1973.

Hibler, W. D., and C. F. Ip (1995), The effect of sea ice rheology on Arctic buoy drift, in *Ice Mechanics*, edited by J. P. Dempsey and Y. D. S. Rajapakse, pp. 255–264, Am. Soc. of Mech. Eng., New York.

Holland, M. M., C. M. Bitz, E. C. Hunke, W. H. Lipscomb, and J. L. Schramm (2006), Influence of the sea ice thickness distribution on polar climate in CCSM3, *Journal of Climate*, *19*(11), 2398–2414, doi:10.1175/JCLI3751.1.

Hopkins, M. A. (1998), Four stages of pressure ridging, *Journal of Geophysical Research*, *103*(C10), 21,883, doi:10.1029/98JC01257.

Hopkins, M. A., and A. S. Thorndike (2006), Floe formation in Arctic sea ice, *Journal of Geophysical Research: Oceans*, *111*(11), 1–9, doi:10.1029/2005JC003352.

Hopkins, M. A., W. D. Hibler, and G. M. Flato (1991), On the Numerical Simulation of the Sea Ice Ridging Process, *Journal of Geophysical Research*, *96*, 4809–4820.

Hunke, E. C. (2014), Sea ice volume and age: Sensitivity to physical parameterizations and thickness resolution in the CICE sea ice model, *Ocean Modelling*, *82*, 45–59, doi: 10.1016/j.ocemod.2014.08.001.

Hunke, E. C., W. H. Lipscomb, and A. K. Turner (2011), Sea-ice models for climate study: Retrospective and new directions, *Journal of Glaciology*, *56*(200), 1162–1172, doi:10.3189/002214311796406095.

Kauker, F., T. Kaminski, R. Ricker, L. Toudal-Pedersen, G. Dybkjaer, C. Melsheimer, S. Eastwood, H. Sumata, M. Karcher, and R. Gerdes (2015), Seasonal sea ice predictions for the Arctic based on assimilation of remotely sensed observations, *The Cryosphere Discussions*, *9*(5), 5521–5554, doi:10.5194/tcd-9-5521-2015.

Kimura, N., A. Nishimura, Y. Tanaka, and H. Yamaguchi (2013), Influence of winter sea-ice motion on summer ice cover in the Arctic, *Polar Research*, *32*, 1–8, doi: 10.3402/polar.v32i0.20193.

Komuro, Y., and T. Suzuki (2013), Impact of subgrid-scale ice thickness distribution on heat flux on and through sea ice, *Ocean Modelling*, *71*, 13–25, doi: 10.1016/j.ocemod.2012.08.004.

Komuro, Y., T. Suzuki, T. T. Sakamoto, H. Hasumi, M. Ishii, M. Watanabe, T. Nozawa, T. Yokohata, T. Nishimura, K. Ogochi, S. Emori, and M. Kimoto (2012), Sea-Ice in Twentieth-Century Simulations by New MIROC Coupled Models: A Comparison between Models with High Resolution and with Ice Thickness Distribution, *Journal of the Meteorological Society of Japan*, *90*(A), 213–232.

Kwok, R., and G. F. Cunningham (2008), ICESat over Arctic sea ice: Estimation of snow depth and ice thickness, *Journal of Geophysical Research*, *113*(C8), C08,010, doi: 10.1029/2008JC004753.

Lavergne, T., S. Eastwood, Z. Teffah, H. Schyberg, and L.-A. Breivik (2010), Sea ice motion from low-resolution satellite sensors: An alternative method and its validation in the Arctic, *Journal of Geophysical Research*, *115*(C10), C10,032, doi: 10.1029/2009JC005958.

Lemieux, J.-F., and B. Tremblay (2009), Numerical convergence of viscous-plastic sea ice models, *Journal of Geophysical Research*, *114*(C5), C05,009, doi: 10.1029/2008JC005017.

Lindsay, R., and A. Schweiger (2015), Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations, *The Cryosphere*, *9*(1), 269–283, doi: 10.5194/tc-9-269-2015.

Lindsay, R., M. Wensnahan, A. Schweiger, and J. Zhang (2014), Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic, *Journal of Climate*, *27*(7), 2588–2606, doi:10.1175/JCLI-D-13-00014.1.

Lipscomb, W. H. (2001), Remapping the thickness distribution in sea ice models, *Journal of Geophysical Research*, *106*(C7), 13,989, doi:10.1029/2000JC000518.

Lipscomb, W. H., E. C. Hunke, W. Maslowski, and J. Jakacki (2007), Ridging, strength, and stability in high-resolution sea ice models, *Journal of Geophysical Research: Oceans*, *112*, 1–18, doi:10.1029/2005JC003355.

Losch, M., D. Menemenlis, J.-M. Campin, P. Heimbach, and C. Hill (2010), On the formulation of sea-ice models. Part 1: Effects of different solver implementations and parameterizations, *Ocean Modelling*, *33*(1), 129–144.

Maslowski, W., and W. H. Lipscomb (2003), High resolution simulations of Arctic sea ice, 1979 1993, *Polar Research*, *22*(1), 67–74.

Massonnet, F., T. Fichefet, H. Goosse, M. Vancoppenolle, P. Mathiot, and C. König Beatty (2011), On the influence of model physics on simulations of Arctic and Antarctic sea ice, *Cryosphere*, *5*(3), 687–699, doi:10.5194/tc-5-687-2011.

McPhee, M. (1975), Ice-Momentum Transfer for the AIDJEX Ice Model, *AIDJEX Bulletin*, *29*, 93–112.

Menemenlis, D., I. Fukumori, and T. Lee (2005), Using Green's Functions to Calibrate an Ocean General Circulation Model, *Monthly Weather Review*, *133*(5), 1224–1240, doi:10.1175/MWR2912.1.

Menke, W. (2012), *Geophysical Data Analysis: Discrete Inverse Theory*, Elsevier, doi:10.1016/B978-0-12-397160-9.00019-9.

Nguyen, A. T., D. Menemenlis, and R. Kwok (2011), Arctic ice-ocean simulation with optimized model parameters: Approach and assessment, *Journal of Geophysical Research:*

Oceans, *116*(4), 1–18, doi:10.1029/2010JC006573.

Parmerter, R. R., and M. D. Coon (1972), Model of pressure ridge formation in sea ice, *Journal of Geophysical Research*, *77*(33), 6565–6575, doi:10.1029/JC077i033p06565.

Parmerter, R. R., and M. D. Coon (1973), Mechanical Models of Ridging in the Arctic Sea Ice Cover, *AIDJEX bulletin*, *19*.

Pritchard, R. S. (1981), Mechanical Behavior of Pack Ice, in *Mechanical Behaviour of Structured Media*, edited by A. P. S. Selvadurai, pp. 371–405, Elsevier.

Richter-Menge, J. A. (1997), Towards improving the physical basis for ice-dynamics models, *Annals of Glaciology*, *25*, 177–182.

Richter-Menge, J. A., and B. C. Elder (1998), Characteristics of pack ice stress in the Alaskan Beaufort Sea, *Journal of Geophysical Research: Oceans*, *103*(C10), 21,817–21,829, doi:10.1029/98JC01261.

Rothrock, D. A. (1975), The energetics of the plastic deformation of pack ice by ridging, *Journal of Geophysical Research*, *80*(33), 4514–4519, doi:10.1029/JC080i033p04514.

Saha, S., S. Moorthi, H. L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y. T. Hou, H. Y. Chuang, H. M. H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. Van Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. Van Den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J. K. Schemm, W. Ebisuzaki, R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C. Z. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R. W. Reynolds, G. Rutledge, and M. Goldberg (2010), The NCEP climate forecast system reanalysis, *Bulletin of the American Meteorological Society*, *91*(August), 1015–1057, doi:10.1175/2010BAMS3001.1.

Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok (2011), Uncertainty in modeled Arctic sea ice volume, *Journal of Geophysical Research*, *116* (September), doi:10.1029/2011JC007084.

Semtner, A. J. J. (1976), A model for the thermodynamic growth of sea ice in numerical investigations of climate, *Journal of Physical Oceanography*, *6*, 379–389.

Steele, M., J. Zhang, D. Rothrock, and H. L. Stern (1997), The force balance of sea ice in a numerical model of the {A}rctic {O}cean, *Journal of Geophysical Research*, *102* (C9), 21,061—-21,079, doi:10.1029/97JC01454.

Stroeve, J. C., A. P. Barrett, M. C. Serreze, and A. Schweiger (2014), Using records from submarine, aircraft and satellite to evaluate climate model simulations of Arctic sea ice thickness, *The Cryosphere*, *8*, 2179–2212, doi:10.5194/tcd-8-2179-2014.

Sumata, H., T. Lavergne, F. Girard-Ardhuin, N. Kimura, M. A. Tschudi, F. Kauker, M. Karcher, and R. Gerdes (2014), An intercomparison of Arctic ice drift products to deduce uncertainty estimates, *Journal of Geophysical Research: Oceans*, *119*, doi:10.1002/2013JC009724.

Sumata, H., R. Kwok, R. Gerdes, F. Kauker, and M. Karcher (2015), Uncertainty of Arctic summer ice drift assessed by high-resolution SAR data, *Journal of Geophysical Research: Oceans*, *120*, 5285–5301, doi:10.1002/2014JC010632.

Thorndike, A. S., D. A. Rothrock, G. A. Maykut, and R. Colony (1975), The Thickness Distribution of Sea Ice, *Journal of Geophysical Research*, *80* (33), 4501, doi:10.1029/JC080i033p04501.

Tucker, W. B., and D. K. Perovich (1992), Stress measurements in drifting pack ice, doi:10.1016/0165-232X(92)90012-J.

Tuhkuri, J. (2002), Laboratory tests on ridging and rafting of ice sheets, *Journal of Geophysical Research*, *107*(C9), 1–14, doi:10.1029/2001JC000848.

Wilchinsky, A. V., and D. L. Feltham (2012), Rheology of Discrete Failure Regimes of Anisotropic Sea Ice, *Journal of Physical Oceanography*, *42*(7), 1065–1082, doi:10.1175/JPO-D-11-0178.1.

Zhang, J., and W. D. Hibler (1997), On an efficient numerical method for modeling sea ice dynamics, *Journal of Geophysical Research*, *102*, 8691–8702.

Zhang, J., and D. A. Rothrock (2001), A Thickness and Enthalpy Distribution Sea-Ice Model, *Journal of Physical Oceanography*, *31*(1), 2986–3001, doi:10.1175/1520-0485(2001)031¡2986:ATAEDS¿2.0.CO;2.

Table 1. Bin limits for ITD configurations

# of categories	bin limits in m									
5	0.0	0.64	1.39	2.47	4.57					
20	0.0	0.16	0.33	0.50	0.67	0.86	1.06	1.28	1.52	1.79
...	2.10	2.46	2.89	3.42	4.06	4.85	5.82	7.01	8.46	10.2

Table 2. Optimized parameters [a]

Parameter		starting values	Baseline	noITD	ITD5R	ITD20R
albedo dry ice	α_{Id}	0.7000	0.71	-	-	-
albedo wet ice	α_{Iw}	0.7060	0.7119	-	-	-
albedo dry snow	α_{Sd}	0.8652	0.8556	-	-	-
albedo wet snow	α_{Sw}	0.8085	0.7903	-	-	-
air drag	$c_{d,a}$	1.14e-3	1.657e-3	-	-	-
water drag	$c_{d,w}$	5.563e-3	6.647e-3	-	-	-
axis ratio	e	2.0	1.523	-	-	-
lead opening	H_0	0.5	0.5649	-	(0.3546)	(0.3292)
ice strength (H79)	P^*	2.264	-	2.299	-	-
ice strength (H79)	C^*	20.0	-	15.92	-	-
ice strength (R75)	C_f	14.0	-	-	13.926	14.07
ridging participation	a^*	0.04	-	-	0.04058	0.04249
ridge shape	μ	4.5	-	-	3.029	3.104

[a] '-' means no change from the last column, values in bracket are from additional optimizations for H_0

Table 3. Cost function values [b]

	Concentration	Thickness	Winter Drift	Summer Drift	Total
Baseline	1.71	0.75	0.52	1.06	4.04
noITD	1.69	0.75	0.50	1.03	**3.97**
ITD5 no tuning	1.84	0.81	1.20	2.00	5.84
ITD5-g1	1.79	0.85	1.06	1.74	5.44
ITD5-g3	1.62	0.75	0.69	1.23	4.28
ITD5-g13	1.67	0.78	0.81	1.39	4.66
ITD5R	1.57	0.72	0.56	1.20	**4.05**
ITD5R-H0	1.49	0.79	0.54	1.22	4.03
ITD20 no tuning	1.91	1.17	0.88	1.56	5.53
ITD20R	1.71	0.90	0.45	1.09	**4.15**
ITD20R-H0	1.63	0.87	0.42	1.11	4.04
ITD5H	1.57	0.63	0.45	0.95	**3.59**
ITD20H	1.77	0.61	0.46	0.91	**3.76**

[b] Experiment names as defined in Table 4

Table 4. Optimized Runs

	initiated from	optimized parameters
Baseline	[*Nguyen et al.*, 2011]	group 1
ITD5-g1	Baseline	group 1
ITD5-g3	Baseline	group 3
ITD5-g13	Baseline	group 1+3
noITD	Baseline	group 2
ITD5R	Baseline	group 3
ITD20R	Baseline	group 3
ITD5H	ITD5R	group 2 taken from noITD
ITD20H	ITD20R	group 2 taken from noITD
ITD5R-H0	ITD5R	H_0
ITD20R-H0	ITD20R	H_0

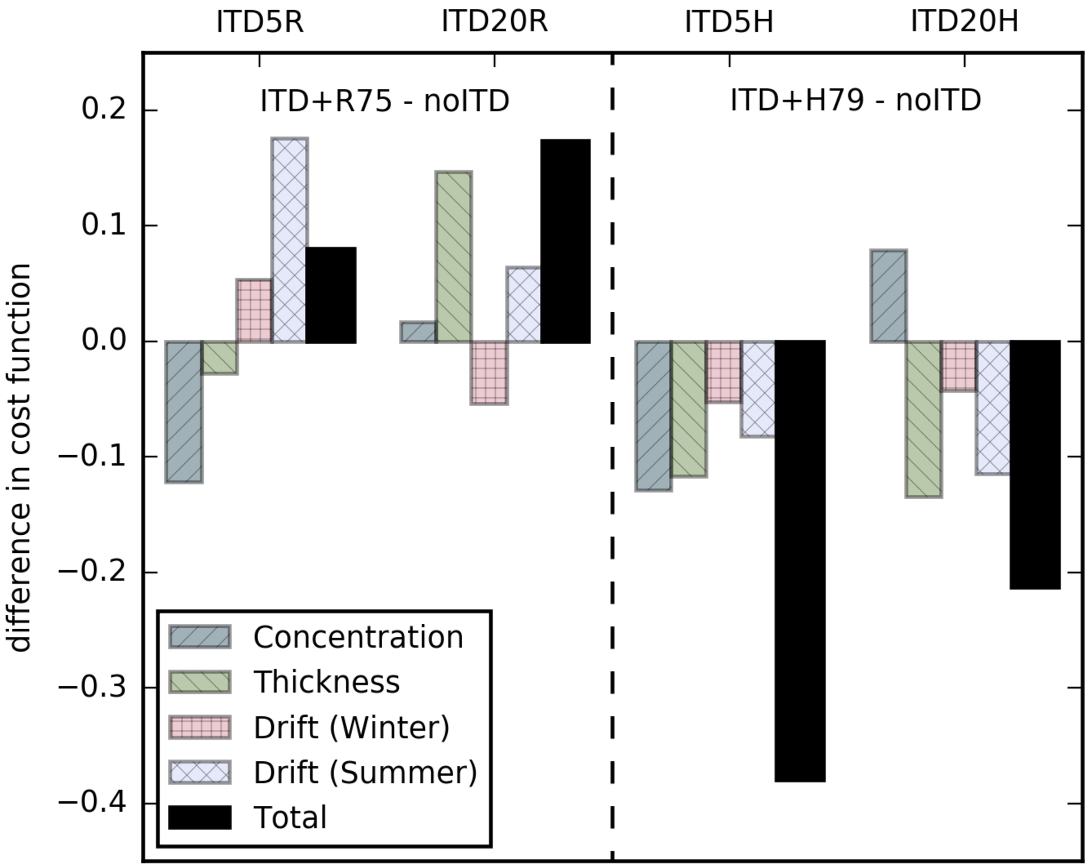

Figure 1. Difference in cost function values (ITD configuration - noITD) between different model configurations with an ITD and noITD. Shown are contributions of single datasets and total values.

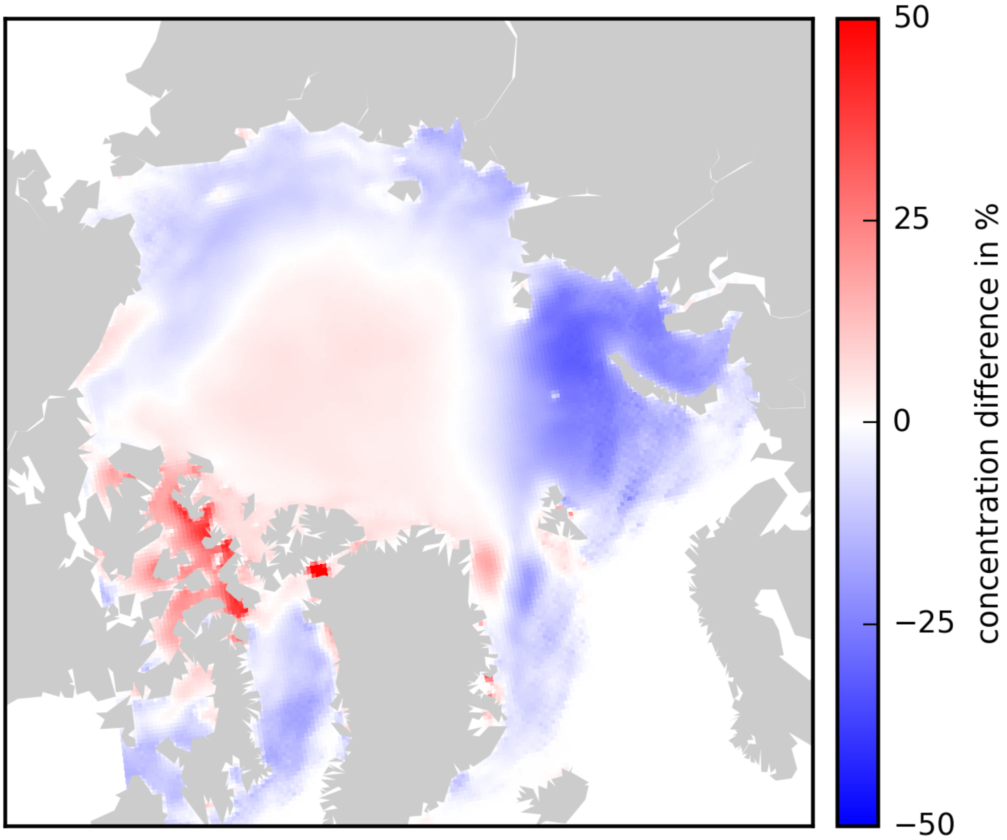

Figure 2. Mean difference in ice concentration (ITD5H - noITD) between an ITD configuration using 5 thickness categories and noITD, both with the H79 strength formulation, in Summer (July to September)

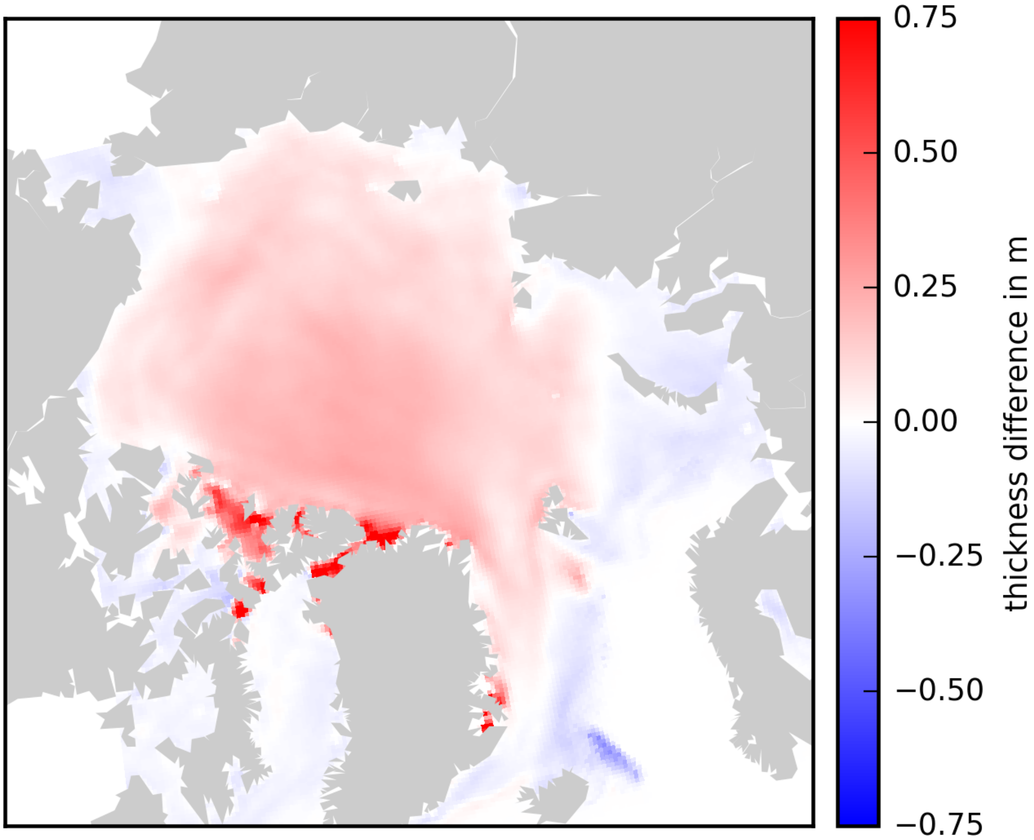

Figure 3. Mean difference in ice thickness H (ITD20H - ITD5H) between ITD configurations with 20 and 5 thickness categories, both using the H79 strength formulation, in Winter (December to May)

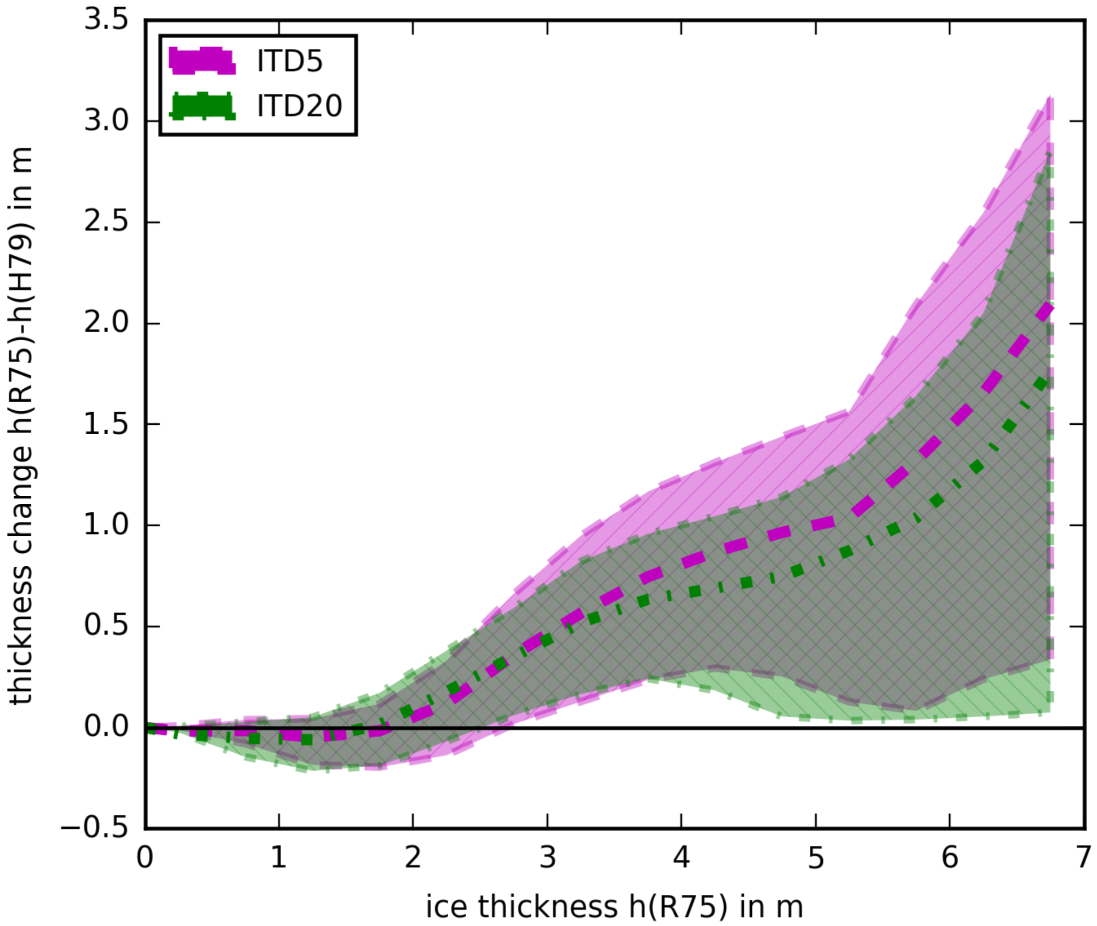

Figure 4. Mean difference in ice thickness (h(R75) - h(H79)) between ITD configurations using R75 and H79 with the same number of thickness categories. The data is binned for ice thickness in the R75 configurations. Purple for ITD5, green for ITD20 with shaded range between 25th and 75th percentile.

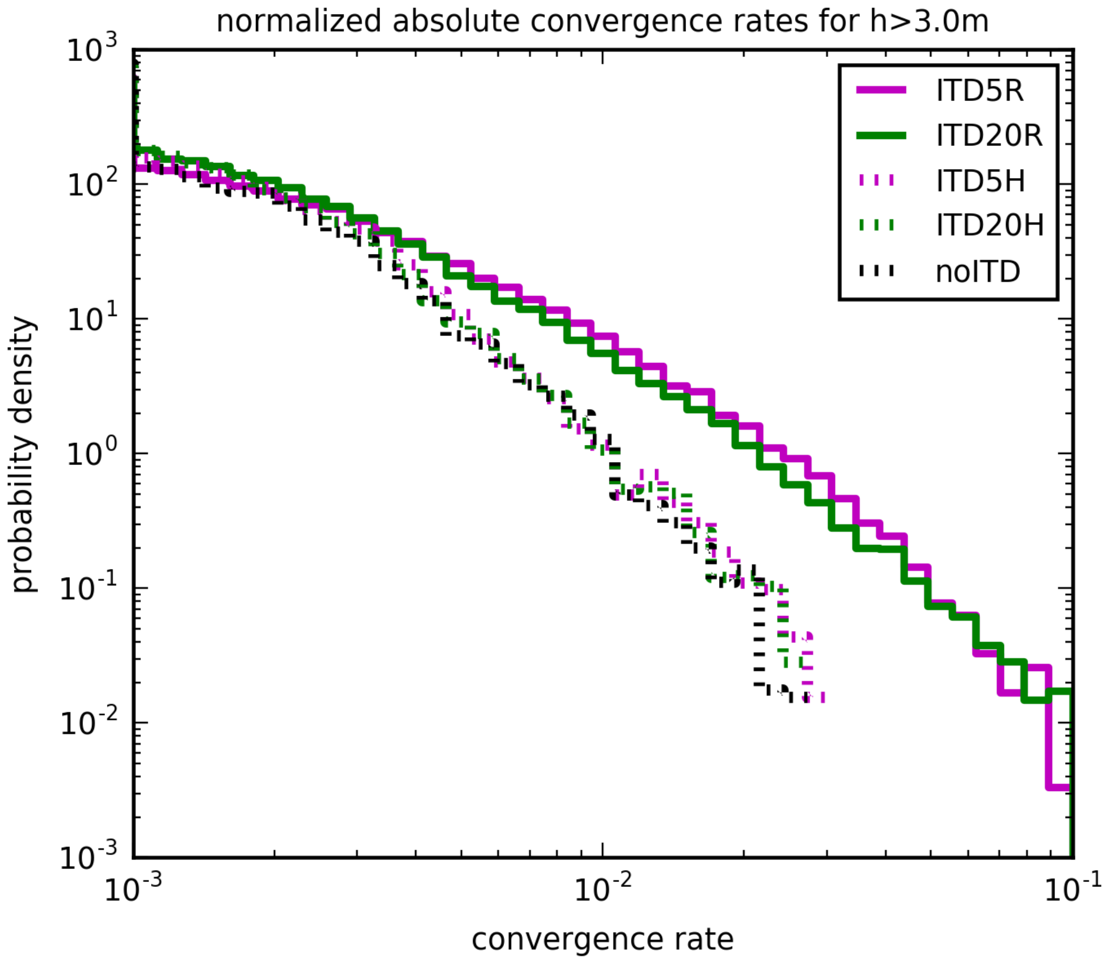

Figure 5. Frequency distribution of absolute convergence rates for configurations ITD5R, ITD20R, ITD5H, ITD20H, noITD; only accounting for ice thicker than 3m.

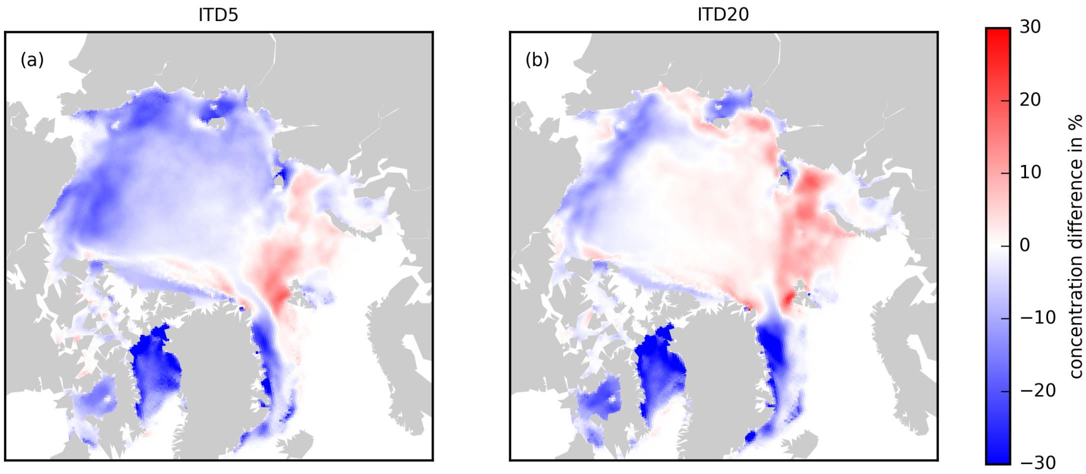

Figure 6. Mean change in August ice concentration (A(H79) - A(R75)) between ITD configurations using H79 and R75 for (a) 5 thickness categories and (b) 20 thickness categories

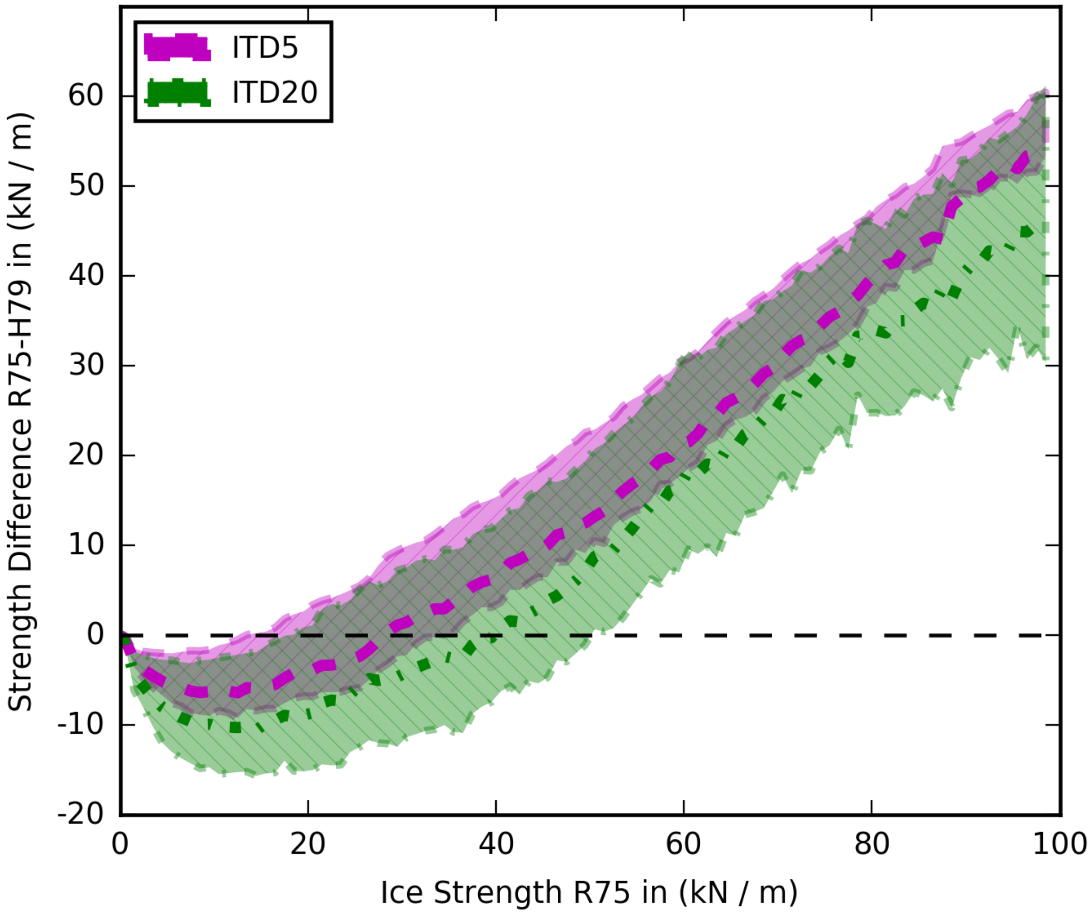

Figure 7. Mean difference in ice strength between R75 and H79 calculated for the same ITD. Differences are evaluated for 5 (magenta) and 20 (green) thickness categories, results are binned for ice strength after R75 with the shaded area between the 25th and 75th percentile.

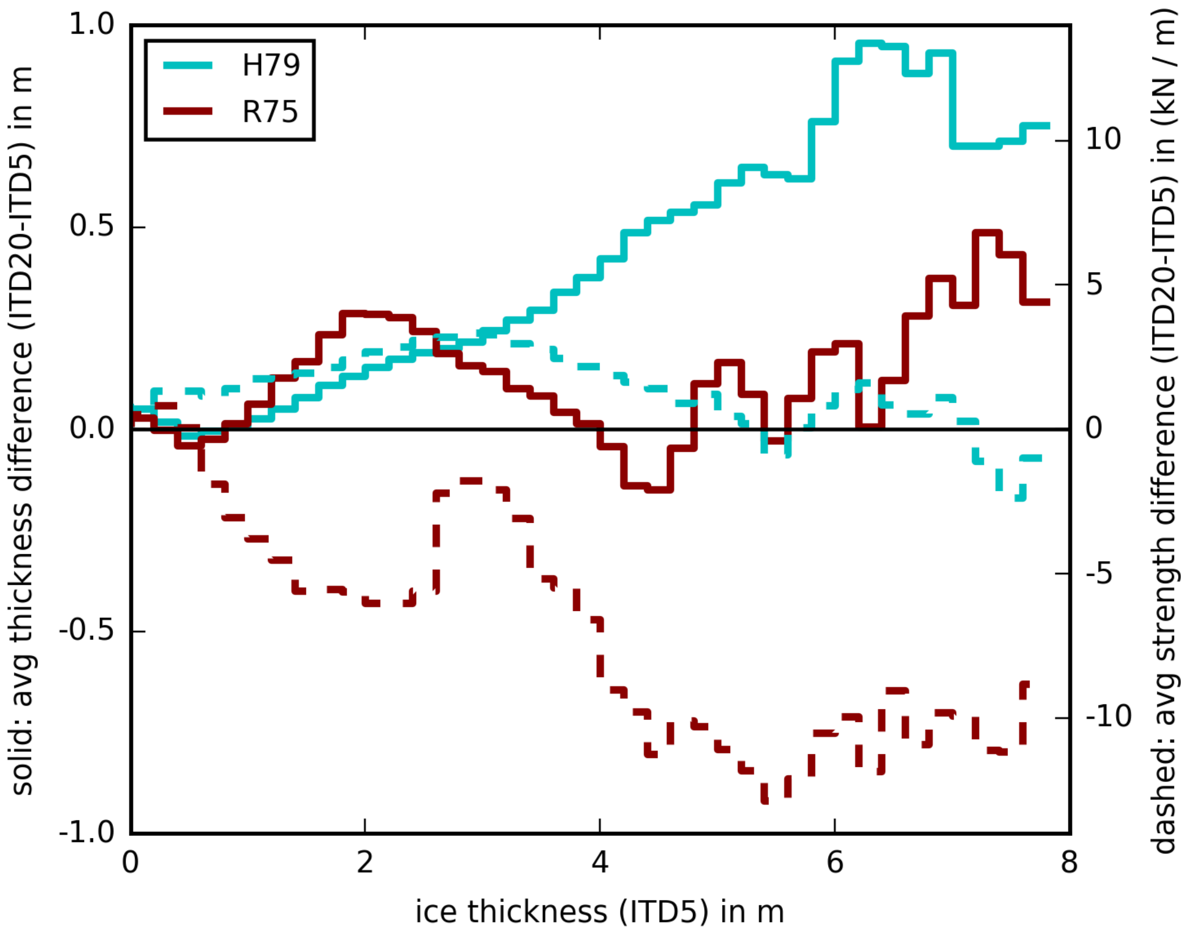

Figure 8. Average difference (ITD20 - ITD5) in ice strength (dashed) and ice thickness (solid) between ITD configurations using 20 and 5 thickness categories evaluated for H79 (cyan) and R75 (red). Differences are evaluated for different ice thicknesses, binned into thickness bins of the ITD5 simulations, as described in section 3.3