Ocean Science

A comparison between gradient descent and stochastic approaches for parameter optimization of a sea ice model

H. Sumata[1], **F. Kauker**[1,2], **R. Gerdes**[1,3], **C. Köberle**[1], and **M. Karcher**[1,2]

[1]Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany
[2]Ocean Atmosphere Systems, Hamburg, Germany
[3]Jacobs University, Bremen, Germany

Correspondence to: H. Sumata (hiroshi.sumata@awi.de)

Abstract. Two types of optimization methods were applied to a parameter optimization problem in a coupled ocean–sea ice model of the Arctic, and applicability and efficiency of the respective methods were examined. One optimization utilizes a finite difference (FD) method based on a traditional gradient descent approach, while the other adopts a micro-genetic algorithm (μGA) as an example of a stochastic approach. The optimizations were performed by minimizing a cost function composed of model–data misfit of ice concentration, ice drift velocity and ice thickness. A series of optimizations were conducted that differ in the model formulation ("smoothed code" versus standard code) with respect to the FD method and in the population size and number of possibilities with respect to the μGA method. The FD method fails to estimate optimal parameters due to the ill-shaped nature of the cost function caused by the strong non-linearity of the system, whereas the genetic algorithms can effectively estimate near optimal parameters. The results of the study indicate that the sophisticated stochastic approach (μGA) is of practical use for parameter optimization of a coupled ocean–sea ice model with a medium-sized horizontal resolution of 50 km × 50 km as used in this study.

1 Introduction

Sea ice plays an important role in shaping the climate system in the Arctic Ocean by altering heat, momentum and material exchanges between the atmosphere and ocean (e.g., Wadhams, 2000; McPhee, 2008; Thomas and Dieckmann, 2009). Development of a sea ice model is thus of great significance not only for understanding the sea ice physics itself but also for understanding the Arctic climate system and its linkage to the global climate. Comprehensive, large-scale sea ice models have existed for more than 3 decades and have provided various insights regarding sea ice and its role in the Arctic climate system (e.g., Hibler, 1979; Hibler and Bryan, 1987; Zhang and Hibler, 1997; Hunke and Dukowicz, 1997). However, even current sea ice models differ in the simulated ice properties and also show pronounced biases compared to observations (Rothrock et al., 2003; Gerdes and Köberle, 2007; Johnson et al., 2007, 2012; Martin and Gerdes, 2007; Eisenman et al., 2007). In order to improve simulated ice properties, explorations of suitable parameterization of dynamic and thermodynamic processes of sea ice (Shine and Henderson-Sellers, 1985; Lipscomb et al., 2007) as well as parameterization regarding atmosphere–ice–ocean fluxes (Lüpkes and Birnbaum, 2005; Lu et al., 2011) are still under way. Such studies are always accompanied by sensitivity studies with respect to newly introduced parameters or an estimation of an optimal parameter set. Particularly, an estimation of an optimal parameter set relevant to respective model configurations is nontrivial work for simulating realistic sea ice fields by a model (Miller et al., 2006; Kim et al., 2006; Nguyen et al., 2011). In this study we explore suitable and effective methods for parameter optimization of coupled ocean–sea ice models, which can be applied to any kind of similar models with relatively small programming effort.

To find an optimal parameter for a parameterization of a certain physical process, early studies performed sensitivity experiments in which a single parameter was varied at a time and other parameters were fixed (Shine and Henderson-

Sellers, 1985; Ledley 1991a, b; Holland et al., 1993). However, Chapman et al. (1994) reported an interdependency of parameter sensitivity and thus the necessity of multivariate sensitivity experiments. In order to find an optimal parameter set in multi-dimensional parameter space, Harder and Fischer (1999) and Miller et al. (2006) performed multivariate sensitivity experiments with a sea ice model. Harder and Fischer (1999) optimized atmospheric and oceanic drag coefficient and ice strength to minimize the misfit between their sea ice model and observations. Similarly, Miller et al. (2006) optimized atmospheric drag coefficient, ice strength parameter and ice albedo simultaneously. In both studies, the parameter space was discretized in a lattice form and combinations of parameters were tested. Although they successfully obtained an optimal parameter set in the three-dimensional parameter space, they had to perform more than 100 experiments even for just 3 parameters. Generally, the number of experiments required for an n-dimensional parameter space increases with the n-th power.

In recent years, more sophisticated approaches for parameter optimization were presented. Kim et al. (2006) applied an automatic differentiation (AD) technique to examine parameter sensitivity of a dynamic thermodynamic sea ice model. In their approach, analytical derivatives of the model with respect to selected parameters were obtained by AD. They used so-called "identical twin experiment" for parameter optimization to demonstrate the performance of their algorithm. It was shown that an AD-based gradient combined with a quasi-Newton search algorithm can effectively retrieve the parameters. Nguyen et al. (2011) presented another sophisticated approach for parameter optimization. They optimized sea ice and ocean model parameters as well as the initial conditions of a coupled ocean–sea ice model with a Green's function approach. Sensitivities of the model with respect to the control parameters were assumed to be linear around a baseline experiment, and then the model Green's function was calculated by perturbation experiments. They obtained a set of parameters, forcing field and initial conditions, which reduces the cost function by 45 %. Other than sea ice models, such methods for model's parameter optimizations can be found in numerous atmospheric and oceanic studies (e.g., Garcia-Gorriz et al., 2003; Menemenlis et al., 2005; Mochizuki et al., 2007; Bocquet, 2012).

Although these approaches provided effective methods to perform a multivariate parameter optimization, problems could arise if the model exhibits a nonlinear response to control parameters, resulting in a complicated shape of the cost function (Evensen and Fario, 1997; Mazzega, 2000). For instance, if the cost function has a multimodal or ill-shaped structure, results of gradient descent methods depend on the initial guess for the parameter set. Generally we cannot exclude the possibility that the cost function has local minima besides the global minimum. In such a situation, one has to perform multiple individual optimizations starting from a variety of initial parameter guesses to find the global minimum.

Another problem may arise from a micro-scale structure of the cost function, because gradient descent approaches can only be reasonably applied if the cost is a smooth function of control parameters. Unfortunately, smoothness of sea ice model's response with regard to its control parameters is not always guaranteed (as will be shown in Sect. 3).

One of the possible solutions to these problems is to apply stochastic algorithms, which perform a random search in the parameter space. Stochastic algorithms, such as simulated annealing or genetic algorithms, are kinds of global optimization algorithms (GOAs) and widely applied to parameter optimization problems in other research fields such as biogeochemical modeling (e.g., Athias et al., 2000; Schartau and Oschlies, 2003; Shigemitsu et al., 2012). Advantages of stochastic approaches are their applicability to multimodal or ill-shaped functions, easy implementation and suitability for parallel computational environment. On the other hand, a serious disadvantage of these approaches is that they generally require huge computational resources as compared with gradient descent approaches for an individual search (Vallino, 2000). This may be one of the reasons why these approaches have not been applied to parameter optimizations for sea ice models coupled with ocean general circulation models (OGCMs), which usually require substantial computational resources. Actually, most of the applications of stochastic approaches for parameter optimizations are found in zero- or one-dimensional model studies (e.g., Carroll, 1996; Vallino, 2000; Athias et al., 2000; Schartau and Oschlies, 2003; Lv et al., 2009), and only a few applications for three-dimensional model studies can be found (Huret et al., 2007).

Difficulties arising from a large computational burden can be overcome by combining parallel processing and a low-computational-cost stochastic approach. Athias et al. (2000) examined the efficiency of parameter optimization by 3 types of GOAs. They reported that a micro-genetic algorithm (μGA) can more efficiently reach a near-optimal solution than other GOAs like simulated annealing (Kirkpatrick et al., 1983) and TRUST (Cetin et al., 1993). The genetic algorithm (GA) is a quasi-stochastic search algorithm to find an optimal solution based on the natural selection of living things (Holland, 1975; Goldberg, 1989). The μGA is a realization with a small computational load by taking advantage of very small population size (Krishnakumar, 1989; Kim et al., 2002). It has been successfully applied to a variety of optimization problems (e.g., Johnson and Abushagur, 1995; Carroll, 1996; Kim et al., 2002; Schartau and Oschlies, 2003), and its advantages compared with the simple GAs were reported by Krishnakumar (1989) and Kim et al. (2002). As will be shown in Sect. 2, the computational load of μGA is quite suitable for multi-processor architecture of recent computers, and a parameter optimization of a coupled ocean–sea ice model can be achieved within a reasonable computational time by using a medium resolution.

The purpose of this study is to provide a simple and systematic method for a parameter optimization of coupled

ocean–sea ice models. In order to provide a suitable optimization method, we introduce a cost function composed of model–data misfit of ice concentration, ice drift velocity and ice thickness. We examine the characteristics of the cost function to specify a requirement necessary for an optimization approach, and then demonstrate parameter optimizations in practice by two types of different approaches: a gradient descent approach and a stochastic approach. As an example of gradient descent approaches, we apply the finite difference method, whereas as an example of stochastic approaches we apply the μGA. By examining applicability and efficiency of the respective approaches together with examinations of the characteristics of the cost function, we provide a useful parameter optimization procedure for modelers working on coupled ocean–sea ice models. To achieve our goal, we demonstrate parameter optimizations with a cost function defined by a 1 yr window. This is too short to estimate proper parameters for a long-term simulation, but long enough to examine the properties of the cost function and the efficiency of the methods. A parameter optimization for a realistic simulation by using a longer time window and examinations of the applicability of the estimated parameters to higher resolution models will be presented in forthcoming papers.

The paper is organized as follows: in Sect. 2 we describe the experiment design, which is composed of a brief introduction of the coupled ocean–sea ice model, sea ice data used in this study, definition of the cost function, a description of two types of optimization methods and a description of optimization experiments. In Sect. 3, properties of the cost function are examined in two-dimensional parameter space, and then results from the two types of optimizations are provided. Conclusions are given in Sect. 4.

2 Experiment design

2.1 Coupled ocean–sea ice model

The coupled ocean–sea ice model used in this study is the North Atlantic/Arctic Ocean Sea Ice Model (NAOSIM) developed at Alfred Wegener Institute (AWI; Gerdes et al., 2003; Köberle and Gerdes, 2003; Kauker et al., 2003). The ocean part of NAOSIM is based on the MOM-2 model developed at the Geophysical Fluid Dynamics Laboratory (GFDL; Pacanowski, 1995), while the sea ice part of the model is a dynamic thermodynamic sea ice model with viscous plastic rheology (Hibler, 1979; Harder, 1996). Both parts of the model are coupled following the procedure devised by Hibler and Bryan (1987). The model domain encloses the whole Arctic and the North Atlantic Ocean north of approximately 50° N (Fig. 1), and is formulated on a spherical rotated grid. The geographical North Pole was shifted to 60° E on the Equator to avoid a numerical singularity arising from convergence of meridians. NAOSIM has been successfully applied to a variety of ocean and sea ice studies in the Arc-
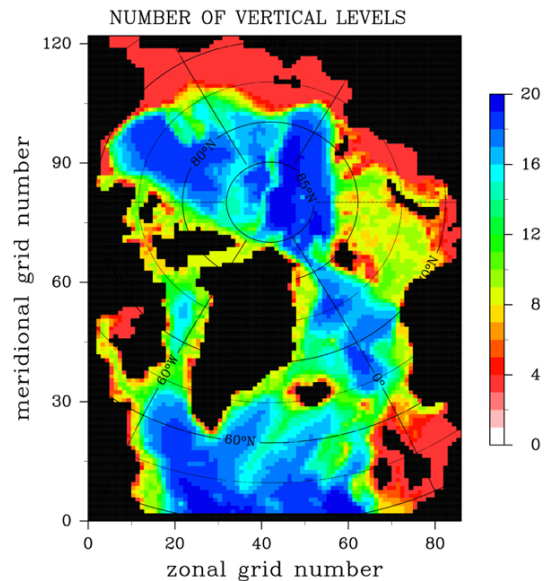

Fig. 1. Bottom topography of the model.

tic Ocean (e.g., Kauker et al., 2003, 2005, 2009; Karcher et al., 2005, 2007, 2011), and more descriptive information about the model configuration can be found in Karcher et al. (2003), Kauker et al. (2003) and studies mentioned above.

In this study we employ a low-resolution version of NAOSIM (Kauker et al., 2009), with a horizontal resolution of $0.5° \times 0.5°$ and 20 levels in the vertical (Table 1). At the southern boundary of the model domain, an open boundary condition has been implemented along the Atlantic sector following Stevens (1991), while in the Pacific sector Bering Strait is treated as a closed wall. At the open boundary of the Atlantic sector, temperature and salinity at inflow points are restored toward PHC (Polar science center Hydrographic Climatology, Steele et al., 2001), and barotropic velocities normal to the boundary are specified from a model version covering the entire Arctic and Atlantic Ocean north of 20° S (Köberle and Gerdes, 2003). The model is driven by daily atmospheric forcing from 1948 to 2003 (NCEP/NCAR reanalysis, Kalnay et al., 1996) starting from temperature and salinity fields given by the PHC climatology and 100 % ice concentration with 2 m thickness in regions where the sea surface temperature falls below the freezing point of sea water. The initial model fields for the parameter optimization window are taken from 1 January 2003, and 1 yr integration window forced by daily NCEP forcing from 1 January to 31 December 2003 is used for parameter optimization.

For parameter optimizations by a gradient descent approach, we make use of two types of model codes. One is the standard model code, and the other is a "smoothed" model code (hereafter referred to as standard- and smth-code, respectively). As will be shown in Sect. 3, responses of modeled sea ice fields to its control parameters do not provide a tractably smooth function for a local gradient estimation.

Table 1. Vertical grid spacing of the model.

Layer number	Layer thickness [m]	Bottom of the layer [m]
1	20.0	−20.0
2	20.0	−40.0
3	20.0	−60.0
4	20.0	−80.0
5	20.0	−100.0
6	21.6	−121.6
7	34.4	−156.0
8	59.3	−215.3
9	95.3	−310.6
10	140.9	−451.5
11	194.0	−645.5
12	252.3	−897.9
13	313.3	−1211.2
14	374.3	−1585.5
15	432.6	−2018.2
16	485.8	−2503.9
17	531.3	−3035.3
18	567.4	−3602.6
19	592.3	−4194.9
20	605.1	−4800.0

Possible reasons are the parameterization of certain physical process and the model's coarse discretization in space and time. The smth-code of NAOSIM was developed for its 4DVar data assimilation system (Kauker et al., 2009) and is used here to mitigate the problems of local gradient estimation arising from the standard-code. In the smth-code, Fortran statements such as "if", "max(.)", "min(.)" and "abs(.)" in the code, which potentially cause discontinuous model behaviors, were replaced by a continuous function such as "atan(.)" to smooth the local structure of the cost function. Although the smth-code slightly modifies the modeled ice and ocean fields, the differences of the simulated fields between the standard- and smth-code are acceptable for the present purposes. Differences of the cost function between the standard- and smth-code will be examined in Sect. 3.

2.2 Data

To evaluate modeled sea ice fields, we make use of 3 types of sea ice data sets obtained from satellite observations (i.e., ice concentration, ice drift velocity and ice thickness). With a basin-wide spatial coverage, these satellite data are suitable to measure model–data misfit and have been applied for parameter optimization of a sea ice model (e.g., Miller et al., 2006) as well as a coupled ocean–sea ice model (e.g., Nguyen et al., 2011).

For sea ice concentration, we use preprocessed sea ice concentration data set of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea Ice Satellite Application Facility (OSI SAF). For the data period in this study, the original data were measured by the Special Sensor Microwave/Imager (SSM/I) and processed following the algorithms described in Eastwood et al. (2010). Here we use the product OSI-409 (available at ftp://saf.met.no/reprocessed/ice/conc/v1/), which contains daily mean ice concentration on a polar stereographic grid with a horizontal resolution of 10 km, covering the entire Arctic Ocean except near the North Pole. We processed the original OSI-409 data set into monthly mean data on the model grid to facilitate model–data comparison. Only the original data whose status flag guarantees its reliability were used. The monthly mean values were defined at points where the number of valid data exceeds at least 30 % of the number of days of the respective month. For the data projection from the data grid to the model grid, we simply calculated the arithmetical mean of valid data contained in each model grid cell. Each grid cell generally contains a sufficient number of data points, and interpolation errors can be negligible.

For sea ice drift, we utilize the low-resolution sea ice drift product OSI-405 from EUMETSAT OSI SAF as well. The data used here are a single sensor product measured by the Advanced Microwave Scanning Radiometer of the Earth Observation System (AMSR-E) and processed following the algorithms described in Lavergne and Eastwood (2010). The data set provides information about positions of ice parcels before and after a certain time interval (48 h) as daily files from January to April and from October to December with some data gaps. In the original data set, the initial positions of the parcels are fixed to the grid points defined on the polar stereographic coordinate with 62.5 km mesh, while the position of the parcels after the time interval is provided as ice displacement data. This procedure introduces certain biases. Nevertheless, this is one of the best available estimates of sea ice motion with large spatial and long temporal coverage. In order to use the data for the present model–data comparison, we calculated monthly mean ice drift on the model grid. In this process, we firstly calculated monthly mean displacement of each parcels on a data grid point when observations were available for more than half of each month. We secondly projected the displacement data from the data grid to the model grid and calculated zonal and meridional ice drift in the model coordinate, with a limitation of maximum interpolation distance of 90 km.

In addition to the above data sets, we also use basin-wide ice thickness data provided by Kwok et al. (2009). The data set is composed of 10 campaigns of Ice, Cloud and land Elevation Satellite (ICESat) from 2003 to 2008 on a polar stereographic grid with a horizontal resolution of 25 km (available at http://rkwok.jpl.nasa.gov/icesat/download.html). Ice thickness is estimated by a method described in Kwok et al. (2007) and Kwok and Cunningham (2008) from total (sea ice plus snow) freeboard measured by a laser altimeter on the satellite. In the present study, we utilized the ON3 campaign from 24 September to 18 November 2003 for model–data comparison. We projected the original data onto the model

grid by simply adopting the nearest data point, since the horizontal resolution of the original data is finer than that of the model.

2.3 Cost function

The cost function measuring the model–data misfit is defined by a combination of 3 types of sea ice data mentioned above and an additional penalty term. The total cost function, J, is given by

$$J = \sum_{k=1}^{3} \frac{J_k}{N_k} + P, \tag{1}$$

where $k = 1$, 2 and 3 correspond to ice concentration, ice drift velocity and ice thickness observations, respectively; J_k represents the contribution from respective component; N_k is the number of observational data for the k-th component; P is a penalty term introduced for a gradient descent approach and will be defined later. To reduce the cost from the respective components simultaneously, we organize the cost function so that the contributions from the respective sea ice properties have the same order of magnitude.

For this purpose, J_k is divided by the number of respective observations, N_k, since the numbers of the respective observations significantly differ from each other ($N_1 = 50\,730$ (ice concentration); $N_2 = 14\,295$ (ice drift velocity); $N_3 = 2123$ (ice thickness)). Together with the observational uncertainties defined later, this normalization makes it possible to evaluate the contribution from the respective components in the same order of magnitude.

We measure each component of the cost function by the squared L^2 norm of model–data misfit weighted by the uncertainties of the observations:

$$J_k = [\boldsymbol{d} - \boldsymbol{G}(\boldsymbol{m})]^T \mathbf{W}[\boldsymbol{d} - \boldsymbol{G}(\boldsymbol{m})], \tag{2}$$

where $\boldsymbol{d} = [d_1, d_2, \dots d_N]^T$ represents the observational data; $\boldsymbol{m} = [m_1, m_2, \dots m_M]^T$ is the control parameter set to be optimized; $\boldsymbol{G}(\boldsymbol{m}) = [G_1(\boldsymbol{m}), G_2(\boldsymbol{m}), \dots G_N(\boldsymbol{m})]^T$ is the convolution of measurement function with the full model dynamics (i.e., $G_i(\boldsymbol{m})$ gives the model's counterpart to the observational data, d_i); $\mathbf{W}$ is the weighting matrix to take uncertainties of the observed data into account. In the present experiments, we only consider the diagonal elements of $\mathbf{W}$ defined by $W_i = \sigma_k^{-2}$, where σ_k is the uncertainty of the respective observations of k-th component. We assume provisional values, $\sigma_1 = 5\,\%$ for ice concentration, $\sigma_2 = 1\,\mathrm{cm\,s^{-1}}$ for ice drift velocity and $\sigma_3 = 50\,\mathrm{cm}$ for ice thickness. These uncertainty values were chosen so as to make the costs associated with respective components have the same order of magnitude and, therefore, contribute to the total cost function to the same extent. An exact evaluation of uncertainties of the merged data and providing an exact form of the weighting matrix is nontrivial and would be quite time-consuming as well as digress from the subject of the present study.

Therefore, we simply assumed the provisional uncertainty values for ice concentration and ice drift velocity, whereas we adopted the uncertainty values provided by Kwok et al. (2009) for ice thickness data, since the thickness data are provided as a time average of each campaign period, and we did not process the data for temporal average.

As control parameters $\boldsymbol{m}$ for the optimization, we selected 7 model parameters from different physical processes as listed in Table 2 (i.e., $M = 7$ in the present case). Two of the parameters were taken from momentum transfer processes among atmosphere, ice and ocean, while others were taken from dynamic and thermodynamic processes of the sea ice model. Atmospheric and oceanic drag coefficients (*cdwin* and *cdwat*) are important tuning parameters for coupled ocean–sea ice models, and a number of studies have focused on obtaining appropriate values for realistic sea ice simulation (e.g., Holland et al., 1993; Chapman et al., 1994; Harder and Fischer, 1999). The present model employs a simple quadratic drag formulation and does not take the atmospheric surface stability into account (i.e., neutral conditions are assumed). The empirical ice strength parameter (P^*) is another key parameter controlling dynamic sea ice processes and has been chosen as one of the tuning parameters in a number of studies (e.g., Holland et al., 1993; Harder and Fischer, 1999; Nguyen et al., 2011). We also include the lead closing parameter in the ice compactness equation, h_0, in the set of control parameters. Furthermore, we select latent and sensible heat transfer coefficients (*cdlat* and *cdsens*) and snow and ice albedo values (*albedo*), since these are the key parameters controlling the thermodynamic processes of sea ice. The representative albedo value in Table 2 is set to be equivalent to the albedo of frozen snow in the model, and is related to other albedo values by maintaining the original ratio between the albedo of frozen snow and other albedo values (0.77/0.8 for melting snow, 0.7/0.8 for frozen ice, and 0.68/0.8 for melting ice). It should be mentioned that the choice of parameters is particular to this study and other Hibler (1979) class of models. A different parameterization of sea ice mechanics and thermodynamics (like in Hunke and Lipscomb, 2001) naturally leads to a different set of parameters and thus different optimization results.

We limited the number of control parameters to 7 to maximize the efficiency of the optimization by the gradient descent approach on our computational environment. This limitation comes from the number of available CPUs in a certain batch job class at the computer facility at AWI, and can be relaxed when we perform a parameter optimization for realistic model simulations. We also set bounds to the parameter values within a prescribed upper and lower limits (see $m_i^{\max}$ and $m_i^{\min}$ in Table 2). These limits are provisional values for examining performances of the present two optimization approaches, and also can later be relaxed within the physical constraints (e.g., that the albedo values should not exceed 1.0).

Table 2. Parameters applied for the optimizations. The representative albedo of snow and ice is related to respective albedo values in the manner described in the text.

m_i	Name	Description	Upper limit ($m_i^{\max}$)	Lower limit ($m_i^{\min}$)	Increment for µGA (possibility $= 2^7$)	Increment for µGA (possibility $= 2^{10}$)
m_1	h_0	Lead closing parameter for ice compactness equation [m]	2	0.1	1.496×10^{-2}	1.857×10^{-3}
m_2	P^*	Empirical ice strength parameter [N m^{-2}]	5.0×10^4	5.0×10^3	354.3	43.99
m_3	$cdwin$	Atmospheric drag coefficient	3.0×10^{-3}	5.0×10^{-4}	1.969×10^{-5}	2.444×10^{-6}
m_4	$cdwat$	Oceanic drag coefficient	1.0×10^{-2}	4.0×10^{-3}	4.724×10^{-5}	5.865×10^{-6}
m_5	$cdlat$	Latent heat transfer coefficient	2.5×10^{-3}	1.25×10^{-3}	9.843×10^{-6}	1.222×10^{-6}
m_6	$cdsens$	Sensible heat transfer coefficient	2.5×10^{-3}	1.25×10^{-3}	9.843×10^{-6}	1.222×10^{-6}
m_7	$albedo$	Representative albedo of snow and ice	0.99	0.6	3.071×10^{-3}	3.812×10^{-4}

The penalty term, P, is introduced so that the gradient descent algorithms can keep the estimated parameters within the prescribed range:

$$P = \sum_{i=1}^{M} \left(\frac{m_i - m_i^{\text{central value}}}{m_i^{\text{half range}}} \right)^{20}, \qquad (3)$$

where $m_i^{\text{central value}} = (m_i^{\max} + m_i^{\min})/2$ and $m_i^{\text{half range}} = (m_i^{\max} - m_i^{\min})/2$, respectively. This term rapidly increases the cost function when the estimated parameters exceed their prescribed ranges, while adding negligible cost when the parameters stay within. Although the stochastic approach does not need the penalty term, we applied the term to a couple of series of optimization experiments with a stochastic approach for comparison purposes.

2.4 Gradient descent approach

As a gradient descent approach, we employ a finite difference (FD) method combined with a quasi-Newton search algorithm to find an optimal parameter set. This is a very fundamental and simple method to optimize model parameters. One of the advantages of the method is that no changes to the model code are necessary to evaluate the gradient of the cost function. Therefore, the method can be easily applied to numerical models of all kinds without a special programming effort. Another advantage is that no linearization of the code is done, and one can explore the fully nonlinear cost function space. On the other hand, the computational cost for gradient estimation increases proportional to the number of the control parameters in contrast to sophisticated approaches such as an adjoint. By exploiting recent parallel computational environments, one can apply the method to parameter optimizations for a moderate number of parameters, while it is still impractical to apply the method to optimize initial and/or boundary conditions of the model. A disadvantage common to all gradient descent approaches is the possibility of becoming stuck in one of the local minima of the cost function.

In order to reduce this possibility, experiments with various initial guesses of the parameter set must be performed.

In the FD method, a gradient of the cost function with regard to each control parameter, $\frac{\partial J}{\partial m_i}$, is evaluated at a certain point in the M-dimensional cost function space by a difference of the cost functions divided by an increment:

$$\frac{\partial J}{\partial m_i} = \frac{J(m + \Delta m_i) - J(m)}{\Delta m_i}, \qquad (4)$$

where $\Delta \boldsymbol{m}_i = [0, 0, \ldots, \Delta m_i, \ldots 0]^T$ is an increment of i-th control parameter normalized by the range of each parameter. By performing $M+1$ model runs and cost evaluations, we can obtain a full gradient of the cost function $\frac{\partial J}{\partial m}$ at a certain point. We applied one of the quasi-Newton search algorithms, the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS) by Liu and Nocedal (1989), to reduce the cost function by searching an optimal parameter set in the gradient descent direction. Re-evaluations of the cost and its gradient and an application of the search algorithm are repeatedly performed, until an Euclidean norm of the gradient (normalized by a Euclidean norm of the control vector) falls below a prescribed threshold or the line search routine in the search algorithm is unable to provide further steps, which satisfies the sufficient decrease and curvature conditions. The threshold for the norm of the gradient is set to 1.0×10^{-12}.

Before performing parameter optimizations, we tested the performance of the L-BFGS algorithm by a pseudo-cost function defined by a M-dimensional continuous function with only one global minimum, and confirmed that the algorithm finds the minimum of the function within a computational accuracy after dozens of iterations. Figure 2a shows schematics of the parameter optimization system by the present method. As shown in the figure, we perform $M+1$ model runs and cost evaluations in parallel by exploiting $M+1$ processors. Therefore, the time required for L iterations is equivalent to L model runs plus the cost evaluations by one processor. Since a parallel computational environment has become quite common, the approach is feasible
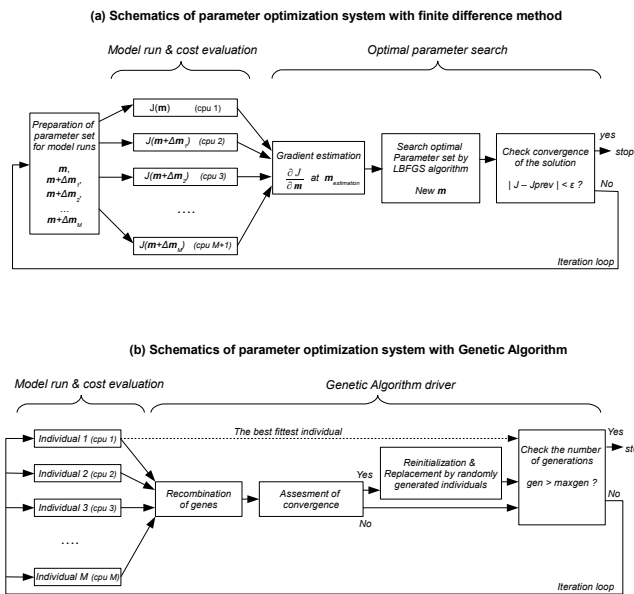
(a) Schematics of parameter optimization system with finite difference method

Model run & cost evaluation   *Optimal parameter search*

(b) Schematics of parameter optimization system with Genetic Algorithm

Model run & cost evaluation   *Genetic Algorithm driver*

Fig. 2. Schematics of a parameter optimization system with **(a)** finite difference method and **(b)** genetic algorithm. See text for description.

for the optimization of a limited number of paramters in a coupled ocean–sea ice model.

Before performing optimization experiments, we also made a survey of the increment, $\Delta \boldsymbol{m}$, for the gradient estimation. If the cost function is a smooth and continuous function of the control parameter set $\boldsymbol{m}$, we can apply an infinitesimal increment for local gradient estimation in Eq. (4). However, as will be shown in Sect. 3, the shape of the cost function given by the coupled ocean–sea ice model is not smooth. It has a spiky micro-scale structure probably due to some parameterizations adopted in the model and the model's discretization in space and time. Although this situation is to some extent alleviated by adopting the smth-code mentioned above, it is not completely solved. A cost function, J, evaluated at a certain point in the parameter space inevitably reflects micro-scale local structure. If we apply a very small increment for a gradient estimation, the gradient represents an inclination of micro-scale unevenness of the function and is not applicable to the minimum search, whereas if we apply a large increment, the gradient cannot represent a local gradient and the accuracy of the search will decline. In such a situation, we have to figure out an appropriate increment size that is sufficiently larger than the width of the micro-scale structure while, at the same time, small enough to capture a local gradient of the function. We performed preliminary optimization experiments with various increment sizes, and evaluated the reduction ratio of the gradient of the function as a measure of an appropriate increment size. If a gradient of the function after an optimization is sufficiently small, the optimized parameter set is close to an extremum of the

function. While the gradient remains large, it is still far from an extremum or the gradient captures micro-scale unevenness of the function due to an inordinately small increment. Therefore, we can say that a smallness of the gradient after an optimization compared to the initial gradient is a necessary condition for being close to an extremum. In the preliminary experiments, we tested various sizes of increment with smth-code and found that an increment with $(m_i^{\max} - m_i^{\min}) \times 10^{-2}$ gives the best reduction ratio of the gradient. We apply this increment value in the following experiments. The adequacy of the increment will be reexamined by two-dimensional survey of the cost function in Sect. 3.

2.5 Stochastic approach

As an example of stochastic approaches, we apply the micro-genetic algorithm (μGA), which is a small population version of the genetic algorithms (GAs). The GAs are global optimization algorithms based on the natural selection of living things (a description of the algorithms can be found in Holland, 1975, and Goldberg, 1989). The advantage of the algorithms is an applicability to an extremum search for ill-shaped or multimodal functions, whereas the disadvantage is a huge computational cost to obtain a solution with a high accuracy. In the algorithms, a single set of the control parameters (parameter vector) is represented by a genotype of an *individual* by encoding the parameter vector to binary bit strings (Fig. 3a), and then a *generation* that is composed of a prescribed number of randomly generated individuals is prepared (Fig. 3b). The generation can be regarded as a pool of various vectors, and the algorithms simulate an evolution of the generation based on a reproductive plan, which consists of selection of individuals, recombination of genes and mutation of individuals. The selection process is conducted by the Darwinian evolutionary principle of "survival of the fittest", which in the present context corresponds to a selection of suitable parameter vector by an evaluation of the cost function (Fig. 3c). The recombination of genes is carried out by an exchange of genes among selected individuals (Fig. 3d), corresponding to a generation of new parameter vector by a random combination of binary bit strings coming from a couple of vectors contained in the pool. The mutation introduces new genic information into the generation, corresponding to an introduction of random seeds of parameter vectors into the pool. Generally, GAs require a population size of $O(10^2)$ to preserve sufficient possibilities for the search. Since the population size is correspondent with the number of model runs required for each generation, it is generally not possible to apply the algorithms to a parameter optimization of coupled ocean–sea ice models in its original form.

The μGA, on the other hand, requires a very small population size regardless of the chromosome length (Goldberg, 1989). The basic strategy of the μGA is to perform a quick search in parameter space by taking advantage of the small population size and to perform an intensive search by
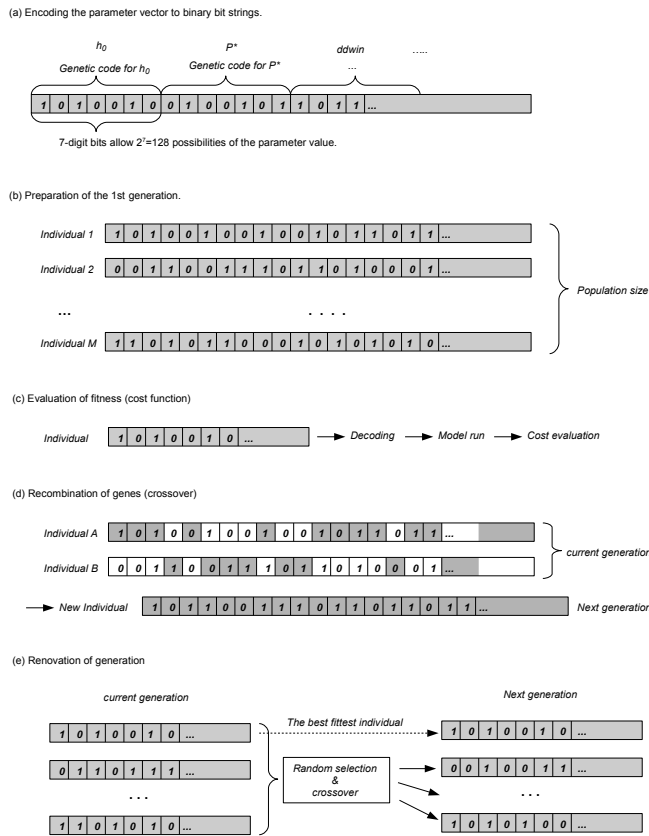
Fig. 3. Schematics of key processes in genetic algorithm: **(a)** encoding the parameter vector to binary bit strings, **(b)** preparation of generation, **(c)** evaluation of fitness of each individual, **(d)** recombination of genes and **(e)** renovation of generation. See text for description.

iterative reinitializations. Technically, the algorithm is composed of 3 processes: a reproduction of individuals, an assessment of convergence and a reintroduction of randomly generated individuals throughout a reinitialization. The reproduction of individuals is achieved by a recombination of genes in the same manner as in the simple GAs with the exception that the best fittest individual of the current generation is reserved and is directly transferred to the next generation without any change (Fig. 3e). After each renewal of generation, the algorithm evaluates convergence of genotype in the generation, and if the diversity of the genes is lower than a certain criterion, all the individuals except the fittest one are replaced by randomly generated new individuals (reinitialization). Since the new information is repeatedly introduced by the reinitialization process, μGA does not need a mutation process. (An example of evolution of estimated parameters is shown in Appendix A.) Advantages of the μGA compared with the simple GAs have been reported in Krishnakumar (1989) and Kim et al. (2002), and examples of application of μGA to various problems can be found in the papers mentioned in the Introduction.

Figure 2b shows schematics of the parameter optimization system with the μGA. We adopted the genetic algorithm driver developed by Carroll (1996) to implement the μGA into the system. As shown in the figure, the number of CPUs necessary for running the system is equivalent to the number of the population size of the generation. The population size required for the μGA is generally less than 10; Goldberg (1989) indicated that a population size of 3 is sufficient for convergence; Coello and Pulido (2001) used a population size of 4; Krishnakumar (1989), Athias et al. (2000) and Kim et al. (2002) used a population size of 5. On the other hand, Schartau and Oschlies (2003) and Shigemitsu et al. (2012), for example, adopted larger population sizes of 13 and 19. In their studies, the population size is set to be the same number as the number of control parameters. Schartau and Oschlies (2003) mentioned that the choice is not mandatory, but was found to perform well in their test experiment. In the present experiment, we adopt population sizes of 5 and 8. The population size of 5 is recommended by a number of former studies, whereas 8 is a number slightly larger than the number of the control parameters. Both of the numbers are quite feasible for recent parallel computational environments. Since the system executes all the model runs in one generation in parallel, the computational time required for an optimization is given by the time for one model run times the number of generations.

The number of generations required for an optimal solution depends on the needed accuracy. Due to the quasi-stochastic nature of the μGA, the convergence of a solution particularly near the optimal solution is quite slow. In addition, since the parameter space is discretized by a prescribed increment and the size of the increment is inversely proportional to the size of the parameter space (or the number of possibilities), a small increment necessary for an accurate solution requires a search in a vast space. Therefore, we have to figure out a practical number of generations as well as a size of an increment in relation to the desired accuracy of a solution. In order to obtain a tentative relation among them, we performed preliminary optimization experiments with a pseudo-function. In these experiments we found that the number of possibilities of 2^7 for each parameter requires at least 400 generations for a solution with 1 % expected error, and 1000 generations for a solution with 0.5 % expected error (see Appendix B for details). It should be noted that since these tentative estimations of errors were obtained from a continuous function with only one minimum, it is not clear whether such an error estimate is applicable to a complicated function with a number of local minima. Nevertheless, from the practical point of view, an optimization experiment with 400 generations is an upper limit of model runs if we intend to apply the method to a parameter optimization with a realistic time window, and therefore we employ these tentative estimations to examine efficiency and applicability of the present method.

Table 3. Five series of optimization experiments with the FD and μGA.

	FD – 1	FD – 2	μGA – 1	μGA – 2	μGA – 3	μGA – 4
Model code	standard	smth	standard	standard	standard	standard
Penalty term	yes	yes	no	yes	no	yes
Number of independent optimization experiments	10	10	10	10	10	10
Number of possible combinations of parameter values	∞	∞	$(2^7)^7$	$(2^7)^7$	$(2^7)^7$	$(2^{10})^7$
Number of population sizes	–	–	5	5	8	5
Number of model runs required for each iteration or generation (except the first generation of the μGA)	8	8	4	4	7	4
Number of total model runs	2536	3232	16 001	16 001	28 001	16 001

2.6 Optimization experiments

Before performing optimization experiments, we conducted two-dimensional complete surveys of the parameter space spanned by h_0 and P^* to demonstrate the nature of the cost function derived from coupled ocean–sea ice model. The surveys were done for both standard- and smth-code with some intensive search in a specific area. All the two-dimensional maps showing the cost function structure are composed of a mesh of 40×40 points obtained from 1600 model runs and cost evaluations.

For the gradient descent approach, we performed a couple of series of optimization experiments with the standard- and smth-code to examine the effect of smoothing of the model code on the efficiency of an optimization (Table 3). The cost function obtained from the smth-code is reevaluated by applying the optimized parameters to the standard-code for comparison purposes. In each series of experiments, 10 independent optimizations starting from different initial parameter sets were performed to see whether the optimized parameter sets converge to certain parameter values (Table 4). The initial parameter sets are generated as follows: the first parameter set is given by the standard parameter setup used in previous studies of NAOSIM, and the second (the third) parameter set is composed of the upper (lower) limits of respective parameter values. The fourth parameter set is composed of a combination of parameter values with their upper (h_0, *cdwat* and *cdsens*) and lower (P^*, *cdwin*, *cdlat* and *albedo*) limits. The fifth to the tenth parameter set are composed of random combinations of selected parameter values, modified in discrete, nearly equidistant steps inside the parameter ranges. It is essential to perform such experiments to assess the applicability of the algorithm to the objective function, because one of the difficulties arising from the approach is missing the global minimum of a function with an ill-shaped structure.

We also performed four series of optimization experiments with the μGA (Table 3). Each series of experiments was composed of 10 independent optimization experiments with different seeds for random number generator in order to statistically assess the efficiency of the algorithm. The first series of experiments (μGA – 1) is conducted with the population size of five and the number of possibilities of $(2^7)^7$. The second series of experiments (μGA – 2) has exactly the same setup as the first one but includes the penalty term in the cost function to facilitate the direct comparison with the optimizations by the FD method. The third (μGA – 3) and fourth (μGA – 4) series of experiments employ different setup parameters in the μGA. The third series uses a population size of eight, whereas the fourth uses a larger number of possibilities of $(2^{10})^7$.

3 Result

3.1 Property of the cost function

Figure 4 shows a two-dimensional cost function map in $h_0 - P^*$ space obtained from the standard-code. The other parameters used for the cost evaluations were fixed to the standard setup values from Table 4. The figure shows that the total cost function reaches its minimum around $h_0 \sim 1.3$ and $P^* \sim 35\,000$ (Fig. 4a) with the standard parameter set. This combination of the h_0 and P^* values does not correspond to minima of the 3 individual components (Fig. 4b–d) of the cost function. The shapes of the J_k/N_k coming from different ice properties significantly differ from each other. For example, around the center of the panels, the response of the ice thickness cost to h_0 variation (Fig. 4b) is opposite to that of the ice concentration cost (Fig. 4c), while at the same time, the ice drift cost is relatively insensitive to h_0 variation (Fig. 4d).

Such property-dependent responses of the cost suggest a couple of important features of the total cost function derived from a combination of different ice properties. The first point

Table 4. Initial parameter sets for optimization experiments with the FD method.

Experiment number	h_0	$P*$	cdwin	cdwat	cdlat	cdsens	albedo
1 (standard setup)	0.5	15 000	2.475×10^{-3}	5.5×10^{-3}	1.75×10^{-3}	1.75×10^{-3}	0.8
2	2.0	50 000	3.0×10^{-3}	1.0×10^{-2}	2.5×10^{-3}	2.5×10^{-3}	0.99
3	0.1	5000	0.5×10^{-3}	4.0×10^{-3}	1.25×10^{-3}	1.25×10^{-3}	0.6
4	2.0	5000	0.5×10^{-3}	1.0×10^{-2}	1.25×10^{-3}	2.5×10^{-3}	0.6
5	0.1	50 000	1.5×10^{-3}	8.0×10^{-3}	2.0×10^{-3}	2.0×10^{-3}	0.7
6	1.0	30 000	2.0×10^{-3}	6.0×10^{-3}	1.5×10^{-3}	1.5×10^{-3}	0.85
7	1.5	40 000	2.5×10^{-3}	8.0×10^{-3}	2.0×10^{-3}	1.25×10^{-3}	0.65
8	0.5	10 000	1.5×10^{-3}	7.0×10^{-3}	1.8×10^{-3}	2.25×10^{-3}	0.75
9	0.75	7500	1.0×10^{-3}	4.0×10^{-3}	1.4×10^{-3}	1.85×10^{-3}	0.95
10	0.25	20 000	0.75×10^{-3}	5.0×10^{-3}	2.4×10^{-3}	2.4×10^{-3}	0.8
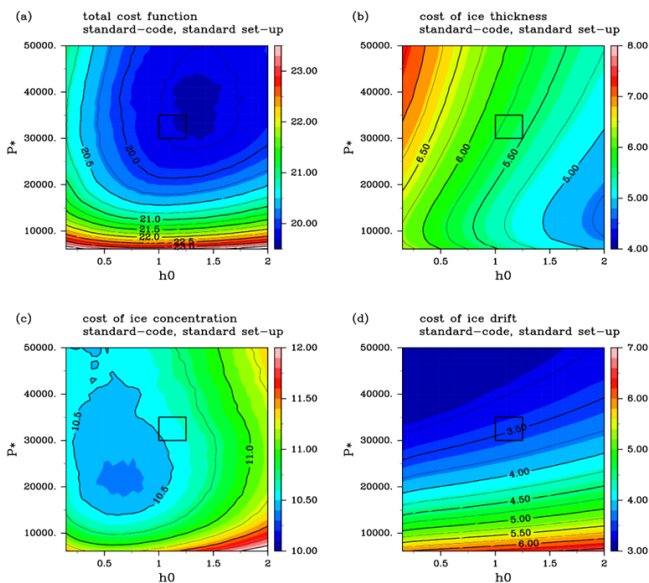
Fig. 4. (a) A two-dimensional structure of the total cost function in $h_0 - P*$ space obtained from the standard-code, and contribution to the cost from **(b)** ice thickness, **(c)** ice concentration and **(d)** ice drift velocity. Note that the same contour interval is adopted in all the panels, whereas the color scale is different. The black solid rectangles around the center of the panel are the area for the magnifications shown in Fig. 5.

Fig. 5. The same as Fig. 4, but the rectangular areas in Fig. 4 are magnified.

is that the function potentially has more than one minimum even if each component of the cost has only one global minimum. The property-dependent responses may be partly due to biased initial sea ice conditions, shortcomings of modeled physics and also partly due to errors in the forcing data. Since a parameter optimization work is inevitably accompanied by such circumstances, the search algorithms for a parameter optimization should be tolerant regarding the existence of local minima. In other words, the gradient descent approaches may have some difficulties finding a global minimum due to the characteristics of the cost function. The second point is
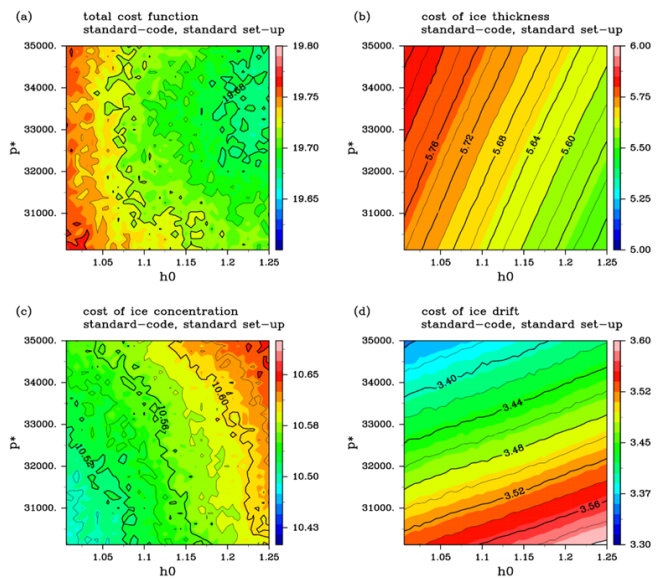
that the shape of the total cost function, and therefore the optimal parameter set corresponding to the cost function, is strongly influenced by relative importance or weightings of respective components. Since the relative importance stems from squared differences between modeled and observed ice properties divided by observational uncertainty, an appropriate choice of respective uncertainties is of significant importance. Although we adopted here the provisional uncertainty values to concentrate on our examination of an applicability and efficacy of the optimization methods, we have to investigate appropriate uncertainties as well as their relative weightings when we perform parameter optimizations for realistic simulations.

Another characteristic of the cost function is its microscale structure. Figure 5 shows a magnification of the black solid rectangles in Fig. 4. The cost function obtained from the
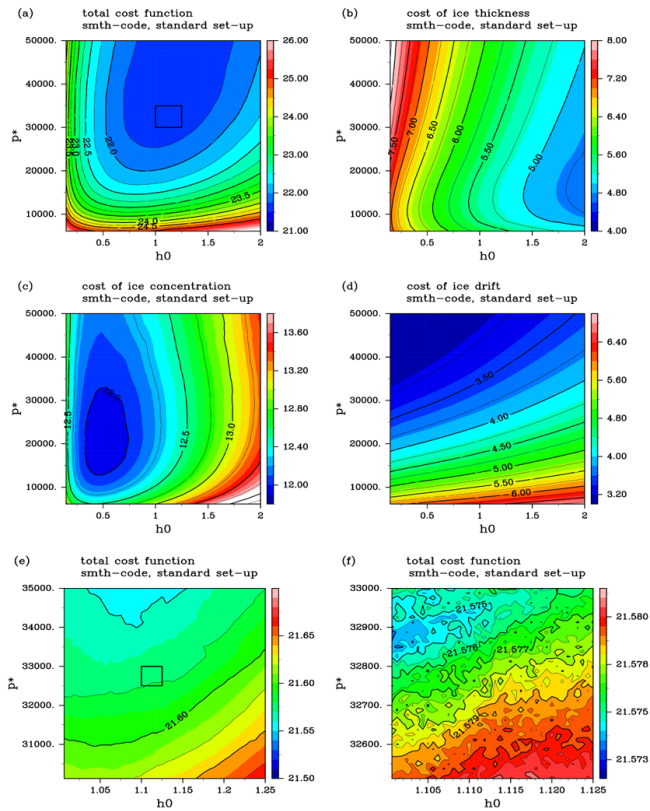
Fig. 6. A two-dimensional structure of the total cost function in $h_0 - P^*$ space obtained from the smth-code, and contribution to the cost from **(b)** ice thickness, **(c)** ice concentration and **(d)** ice drift velocity. **(e)** Magnification of the black solid rectangle in **(a)**. **(f)** Magnification of the black solid rectangle in **(e)**. **(a–d)** utilize the same contour interval but different color scale, whereas **(e)** and **(f)** utilize a different contour interval and color scale.

Fig. 7. Time series of ice concentration cost (blue lines) and ice drift cost (red lines) obtained from the standard setup (solid lines) and from the optimized parameter set (dashed lines). The optimized parameter set is obtained by using finite difference method with the standard-code, starting from the standard setup (see Table 4).

standard-code has a micro-scale uneven structure. Although most of the unevenness on this scale stems from the ice concentration cost, the ice drift and ice thickness cost also have uneven structures if we magnify the map further. Such an uneven structure makes it difficult to accurately estimate a gradient of the function required for the gradient descent approach, and then extremely lowers the accuracy of the solution. Although an application of the smth-code mitigates such a situation to some extent (Fig. 6a, e), the cost function still has an uneven structure on a smaller scale (Fig. 6f). If we estimate possible increments for the FD method from Fig. 6, $\Delta h_0 \sim 0.025$ and $\Delta P^* \sim 500$ are respectively the smallest values that will not be affected by the micro-scale unevenness. These estimations are consistent with the increment values obtained from the preliminary experiments, $(m_i^{max} - m_i^{min}) \times 10^{-2}$ (i.e., $\Delta h_0 = 0.019$ and $\Delta P^* = 450$). It should be kept in mind that although the smth-code does not change the basic responses of the cost function to variations of respective parameters, it increases the cost by ap-
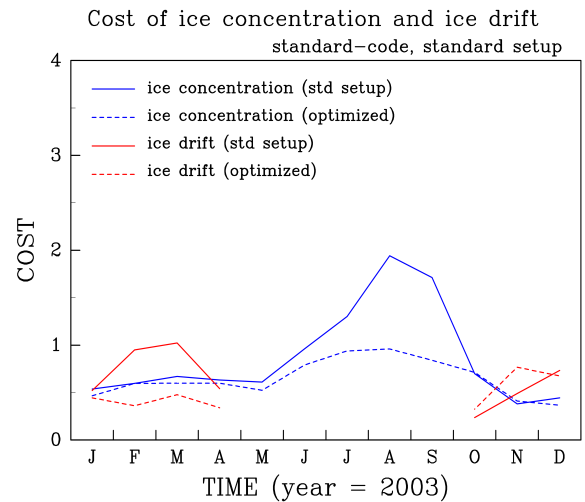
proximately 7–8 % and slightly deforms the shape of the cost function (Fig. 6a–d).

3.2 Gradient descent approach

To assess the ability of the optimization system, we first examine the cost function and corresponding sea ice fields obtained from an optimized parameter set. As an example, Fig. 7 shows a time series of ice concentration cost and ice drift cost before and after an optimization. The optimization was performed by the FD method with the standard-code starting from the standard setup shown in Table 4. The total cost function was reduced by 26.3 % from 20.87 (standard setup) to 15.39 (after the optimization) after 44 iterations, and all of the 3 components of the cost were also reduced (ice concentration cost: 25.6 %, ice drift cost: 24.6 %, ice thickness cost: 28.7 %). As shown in Fig. 7, the system successfully reduced the ice concentration cost in summer, while the cost in winter and spring seasons remained unchanged. In August the broad ice-free area in the Eurasian Arctic found with the standard setup was drastically improved by an application of the optimized parameter set (Fig. 8a–c). As for ice drift velocity, the time series of the cost shows that the cost was reduced in the early months of the year, while the costs in October and November were rather increased (Fig. 7). Figure 8d–f show monthly mean ice drift velocity fields in March, in which the improvement of the modeled ice velocity field is evident. For the ice thickness field, we have only one ICESat campaign in 2003. The ice thickness averaged over the period also shows significant improvement in the Eurasian Arctic (Fig. 8g–i). These results indicate that the
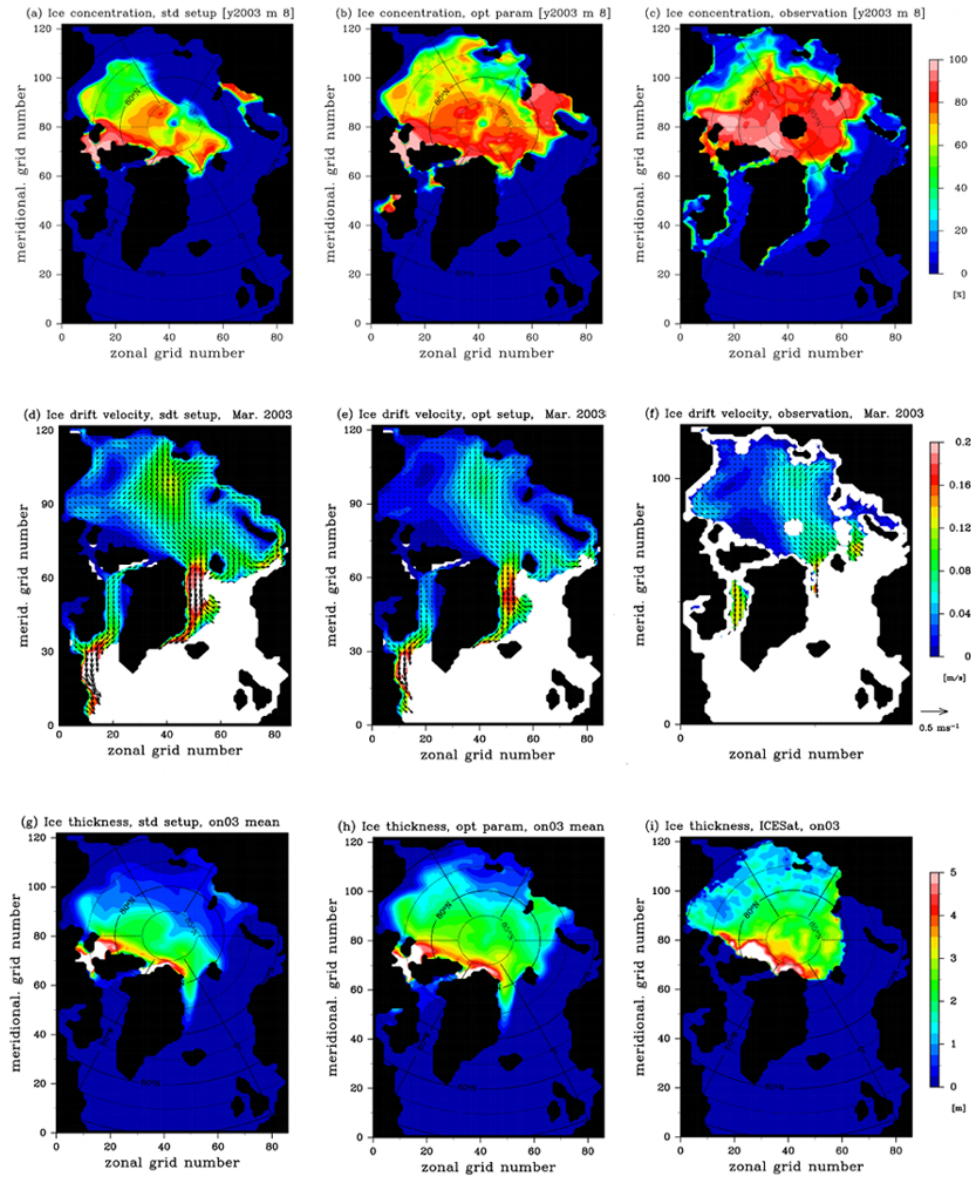
Fig. 8. Modeled and observed **(a–c)** monthly mean sea ice concentration in August 2003, **(d–f)** monthly mean ice drift velocity in March 2003 and **(g–i)** ice thickness in ON3 ICESat campaign period. The panels in the left column are NAOSIM with standard setup, while the panels in the center column are NAOSIM with the optimized parameter set obtained from the finite difference method with the standard-code, starting from the standard setup. The panels in the right column are from satellite observations: **(c)** OSI – 409, **(f)** OSI – 405 and **(i)** ICESat.

system can at least provide a parameter set that gives smaller cost function, although the set might not be optimal.

Figure 9 shows the initial and optimized cost functions and associated parameter values obtained from the 10 independent optimization experiments by the FD method with the standard-code. As shown in Fig. 9a, all the optimizations starting from various parameter sets successfully reduced the cost to similar values as the optimization starting from the standard setup (opt-1). The final costs range from 15.02 to 15.47 (mean value is 15.26). Nevertheless, the estimated parameter values except albedo show quite divergent distribu-

tions despite the fact that they give similar costs (Fig. 9b–h). Possible reasons for the divergent distributions are (1) shortcomings of the FD method associated with an inaccurate gradient estimation caused by the micro-scale unevenness of the cost function and its manifestation particularly in weak-sensitivity regions, and (2) the search algorithm being stuck to one of the local minima of the function. The first possibility comes from the characteristics of the cost function shown in Figs. 4 and 5, and also from the fact that the simulated ice fields corresponding to the estimated parameter sets show quite similar spatial patterns, regardless of the different
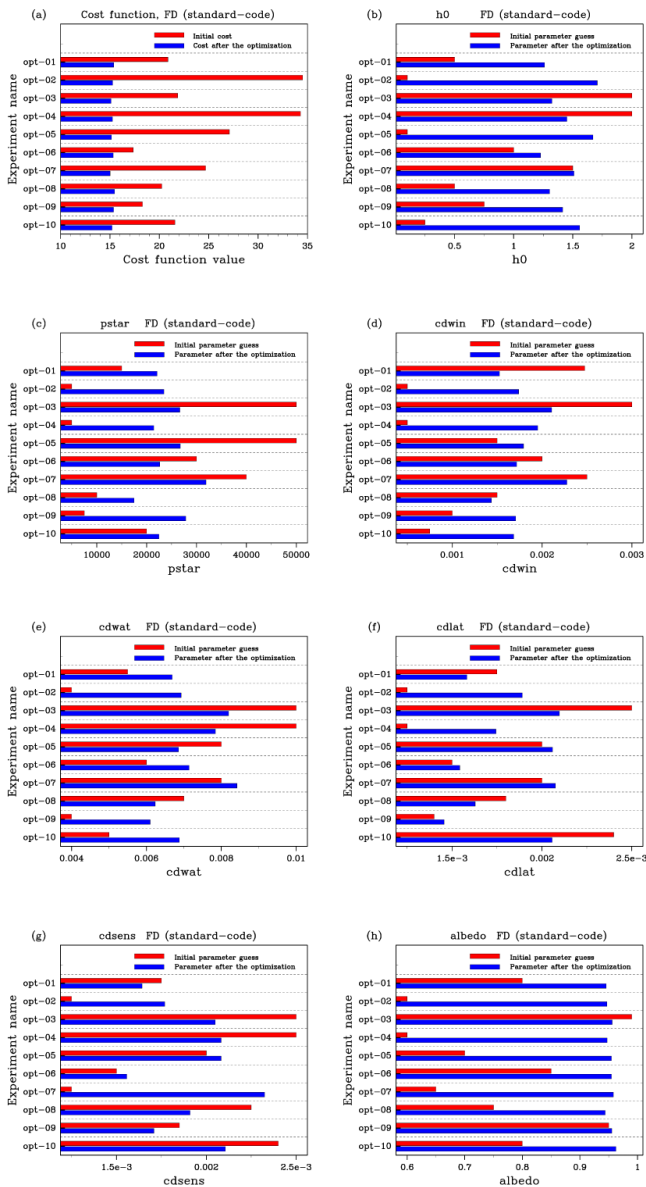
Fig. 9. The initial and optimized cost functions and associated parameter values obtained from 10 independent optimization experiments with the finite difference method using the standard-code. **(a)** Initial (red) and optimized (blue) cost function values. **(b–h)** Initial (red) and optimized (blue) parameter values: **(b)** h_0, **(c)** P^*, **(d)** *cdwin*, **(e)** *cdwat*, **(f)** *cdlat*, **(g)** *cdsens* and **(h)** *albedo*.

parameter values (Fig. 10, the upper panels). The second possibility comes from the fact that some of the estimated parameters tend to have some specific values as can be seen in Fig. 9c–g (e.g., in Fig. 9f, opt-1, 6 and 8 give similar estimated values around 1.6×10^{-3}, whereas opt-3, 5, 7 and 10 give values around 2.1×10^{-3}) . In any case, we can conclude that the FD method with the standard-code can reduce the cost to some extent, but at the same time, has a weakness for estimating parameters with weak sensitivity to the cost.

The application of the smth-code mitigates the divergent property of the estimated parameters to some extent. Figure 11 shows the initial and optimized cost functions and associated parameter values obtained from the 10 independent optimization experiments with the smth-code. Variations of some of the estimated parameters become small compared to the standard-code (Fig. 11b, g and h). Other parameters still vary over a similar range as with that of the standard-code (Fig. 11e and f). The cost function values after the optimizations range from 16.26 to 16.39, and those values corresponding to the same parameter sets but re-evaluated by the standard-code range from 15.26 to 15.59 (mean value is 15.38). These values are slightly larger than those obtained from the optimizations with the standard-code, indicating that the application of the smth-code, as a result, could not provide parameter values, which gives smaller cost than the cost obtained from the standard-code.

3.3 Stochastic approach

In this subsection, we survey applicability and efficiency of the μGA by examining results from μGA − 1, in which the population size is set to 5 and the number of possible parameter values is set to $(2^7)^7$ (see Table 3). Afterwards, we briefly show results from a different setup, μGA − 2, 3 and 4, to make a direct comparison of optimization efficiency between the μGA and the FD method, and also to examine the effects of setup parameters used in the μGA.

The parameter sets obtained from the μGA optimization experiments also succeeded to reduce the cost function and improved the simulated ice fields as summarized in Figs. 12 and 13. The time series of the cost function (Fig. 12) and simulated ice fields (Fig. 13b, e and h) were obtained from a model run for which the mean parameter values obtained from 10 independent optimization experiments by μGA − 1 were applied. The cost function value after the optimization is 14.82, and the reduction rate of the cost compared to the standard setup is 29.0 % (ice concentration cost: 27.5 %, ice drift cost: 28.7 %, ice thickness cost: 32.0 %). As shown in Fig. 12, the reduction of ice concentration cost is emphasized in the summer months similar to the FD optimizations. The costs from January to June are also reduced more than 10 %. For ice drift velocity, the cost decreases in the early months of the year, whereas they increase in October and November, again similar to the FD optimizations. The spatial pattern of the simulated ice fields (Fig. 13b, e and h) also exhibit similar but slightly better structure than those from the FD optimizations (see also Fig. 8b, e and h for comparison).

Figure 14 summarizes the evolution of the cost functions and corresponding parameter values throughout the 10 independent optimization experiments by the μGA (μGA − 1). As shown in Fig. 14a, the cost functions rapidly reduced during the first 100 generations, and the reduction of the cost throughout the following 300 generations is relatively small. After 400 generations, the costs obtained from
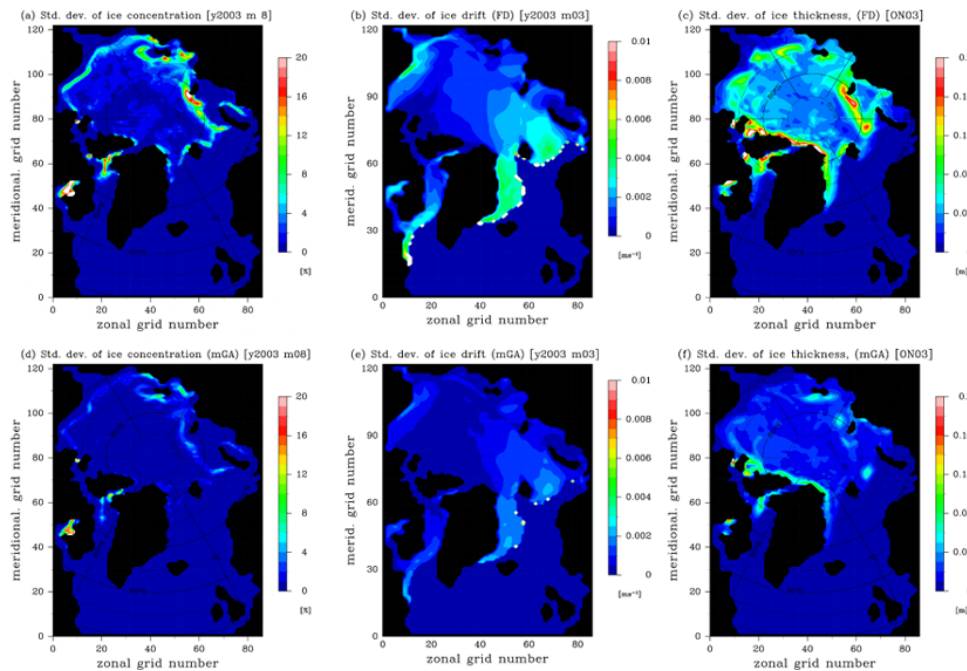
Fig. 10. Standard deviation of the simulated sea ice fields calculated from the 10 model runs. In the respective model runs, the optimized parameter sets obtained from the 10 independent optimization experiments by the finite difference method using the standard-code (the model runs in **a–c**), or those by the μGA – 1 (the model runs in **d–f**) were applied. The standard deviations of (**a** and **d**) monthly mean sea ice concentration in August 2003, (**b** and **e**) monthly mean ice drift velocity in March 2003 and (**c** and **f**) ice thickness in ON03 ICESat campaign period.

the 10 optimization experiments range from 14.80 to 14.83 (mean value is 14.81). The evolution of the associated parameter values, on the other hand, shows slight modifications even after the 100th generation in some cases. The variances of the estimated parameters become nearly constant after 200 generations. (For line plots showing evolution of estimated parameters, see Appendix C.) It should be noted that the estimated parameters are not necessarily applicable to a long-term simulation and/or to assess physical processes of the model, regardless of the reduction of the cost and the convergence of the estimated parameters. This is mainly due to the assimilation window used in our test case being shorter than the spin-up time of the sea ice–ocean system. For example, extremely high albedo values (∼ 0.99) are probably due to the strongly biased ice concentration and thickness found in the original model run (Fig. 14a and g). The algorithm leads to more ice on the Eurasian Basin side by reducing sea ice melt in summer with extremely high albedo values. For realistic parameter estimation, the assimilation window should be at least the spin-up time of the sea ice–ocean system (i.e., at least about 7 yr). This work will be done in a forthcoming paper.

A remarkable point of μGA – 1 optimizations is the small variance of the estimated parameter values even for parameters that varied considerably in the FD method, such as *cdwat* and *cdlat* (Fig. 14; see also Figs. 9 and 11 for comparison).

Figure 15 summaries standard deviations of the estimated parameters obtained from the 10 independent optimization experiments for the respective approaches. μGA – 1 generally provides better convergence of the estimated parameters compared to the FD method. Although the FD method with smth-code provides slightly better convergence than the μGA for *cdsens* and *albedo* values, the standard deviations are quite small in both approaches and the difference between the two approaches is vanishing.

Figure 10d–f show the spatial pattern of standard deviations of the simulated ice fields calculated by the 10 model runs corresponding to the 10 estimated parameter sets by μGA – 1. The figure shows that differences among the simulated ice fields are sufficiently small and generally limited to the marginal ice zone. In addition, the differences are quite small compared to those obtained from the FD method (Fig. 10a–c), indicating better convergence of the simulated ice fields than those obtained from the FD method. The small standard deviations of the simulated ice fields indicate that the corresponding cost functions are closer to the global minimum than those obtained from the FD method.

In order to make a direct comparison of the optimization efficiency between the μGA and the FD methods, we performed an additional series of experiments (μGA – 2), in which the penalty term was included in the cost function. In this series of experiments, all the cost function values after
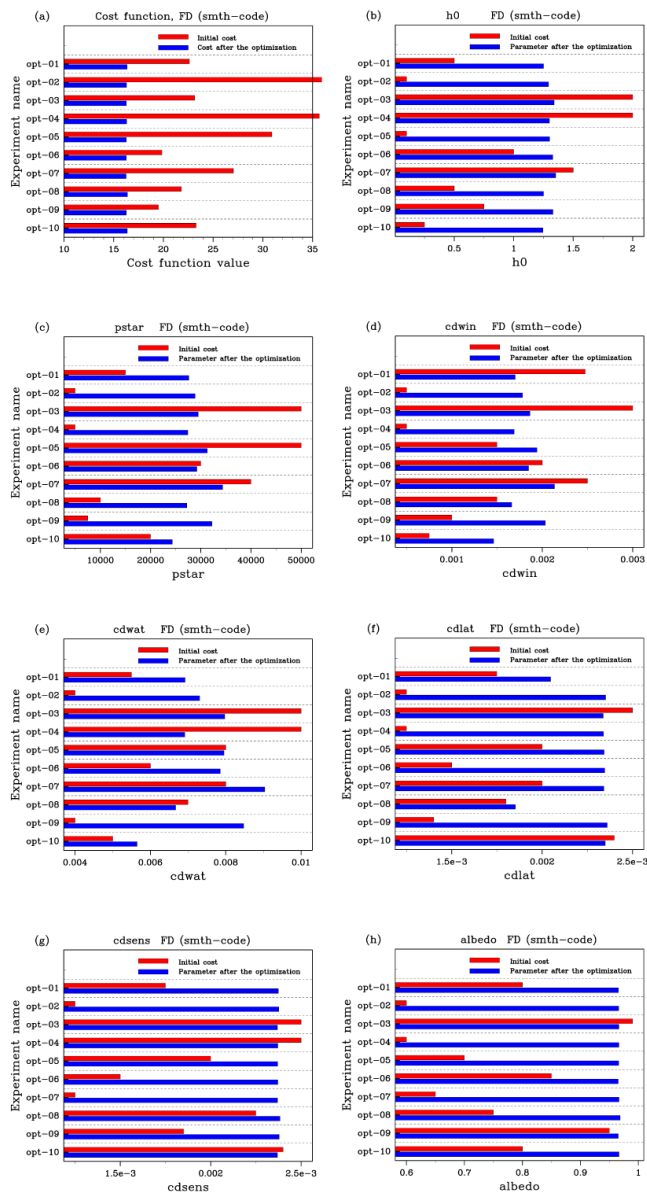
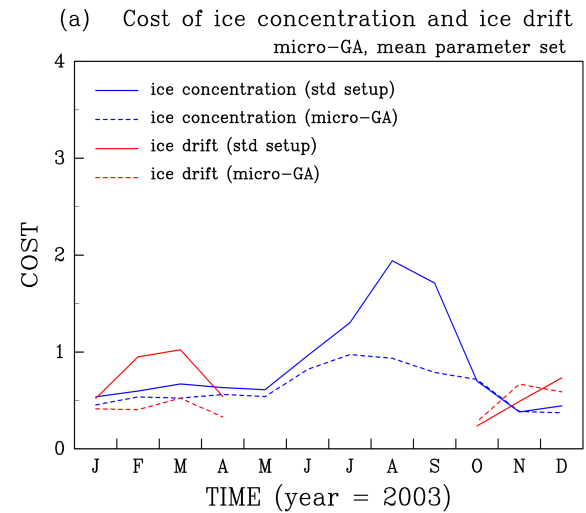**Fig. 11.** The same as Fig. 9 but with the smth-code.

Fig. 12. The same as Fig. 7 but with the μGA optimization experiments (μGA – 1: population size = 5, number of possibilities of each parameter = 2^7).

400th generation), probably because of a reduced number of reinitializations due to increased population size. For a large number of possibilities, the convergence of the solution again decreases, indicating that the larger parameter space worsens the efficiency of the optimization even for an identical cost function. These results indicate that the selection of the setup parameters used in the μGA is important to achieve a fast convergence within a limited number of generations.

4 Summary and conclusion

Two types of optimization methods were applied to a parameter optimization problem of a coupled ocean–sea ice model, and a comparison of the two methods was made to assess an applicability and efficiency of the respective methods. One is a finite difference method based on a gradient descent approach, while the other adopts the stochastic approach of genetic algorithms. To evaluate modeled sea ice properties, a cost function composed of model–data misfit of ice concentration, ice drift velocity and ice thickness was introduced.

An example of a two-dimensional complete survey of the cost function in the parameter space showed that the cost function composed of a combination of different types of observations potentially has more than one minimum value. In addition, the survey also showed that the cost function exhibits a micro-scale uneven structure in parameter space, which prevents the estimation of the gradient of the function with high accuracy. Due to the nature of the cost function, the finite difference (FD) method has difficulties estimating optimal parameters. The estimated parameter values depend on their initial guesses, and the standard deviations of the estimated parameters calculated from 10 independent optimization experiments were quite large for some parameters.
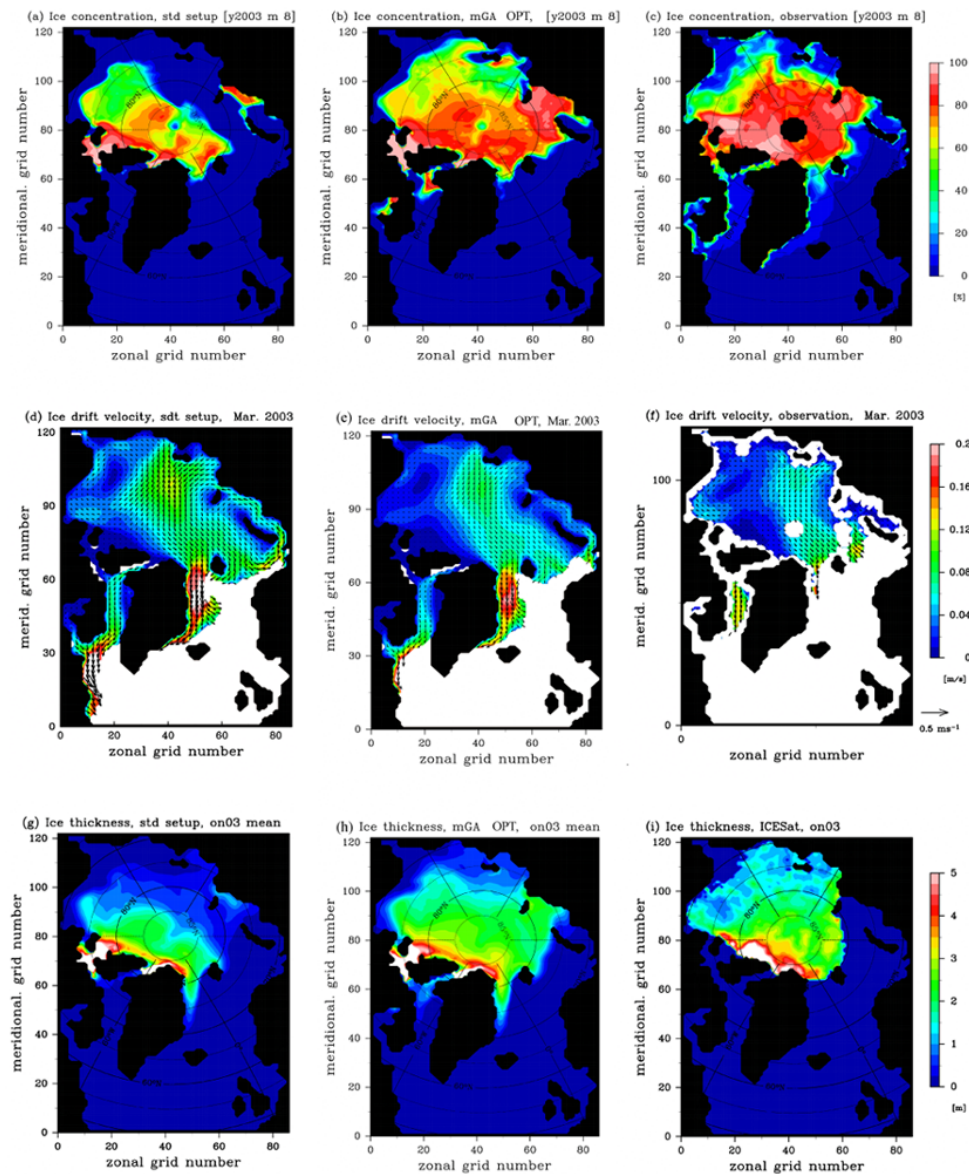
400 generations are smaller than the minimum cost obtained from the FD method, and the standard deviations of the estimated parameters are again satisfactorily small (less than 9 % of the prescribed parameter range), again supporting the advantage of the μGA compared to the FD method.

Further experiments were also performed to examine effects of the setup parameters in the μGA on the efficiency of the optimization. One series of experiments examines the effect of the population size (μGA – 3), and the other series of experiments examines the effect of the number of possible parameter values (μGA – 4). If we increase the population size from 5 to 8, the efficiency of an optimization slightly decreases (the maximum standard deviation of the estimated parameters is 12.3 % of the prescribed parameter range at

Fig. 13. The same as Fig. 8 but with the average of the μGA – 1. The results from the standard setup (the left column) and from satellite observations (the right column) are again shown to facilitate comparisons.

A longer time window for a more realistic parameter optimization would render the cost function more complicated and would make the FD method even harder to apply. The genetic algorithm, on the other hand, provides satisfactory results regardless of a relatively small number of generations used here. The results show that the standard deviations of estimated parameters calculated from 10 independent optimization experiments were less than 6 % of the prescribed range of the respective parameters. Examinations of standard deviations of the simulated ice fields also suggest that the μGA can provide cost functions that are closer to the global minimum than those from the FD method. From these results, we conclude that the μGA is favorable compared to the

FD method for a parameter optimization of coupled ocean–sea ice model of medium resolution.

In a forthcoming paper we will extend the assimilation window to be comparable to the spin-up time of the sea ice model of about 5 to 7 yr to obtain more "physical" optimal parameters. This set of parameters can be applied to higher resolution models ($O \sim 10\,km$). The cost function can be evaluated without large computational costs, and the benefit can be examined although the parameters are not "optimal" for these models. A direct application of the μGA to the $O\,(10\,km)$ model is possible but not achievable with the computer resources currently available to us at AWI .
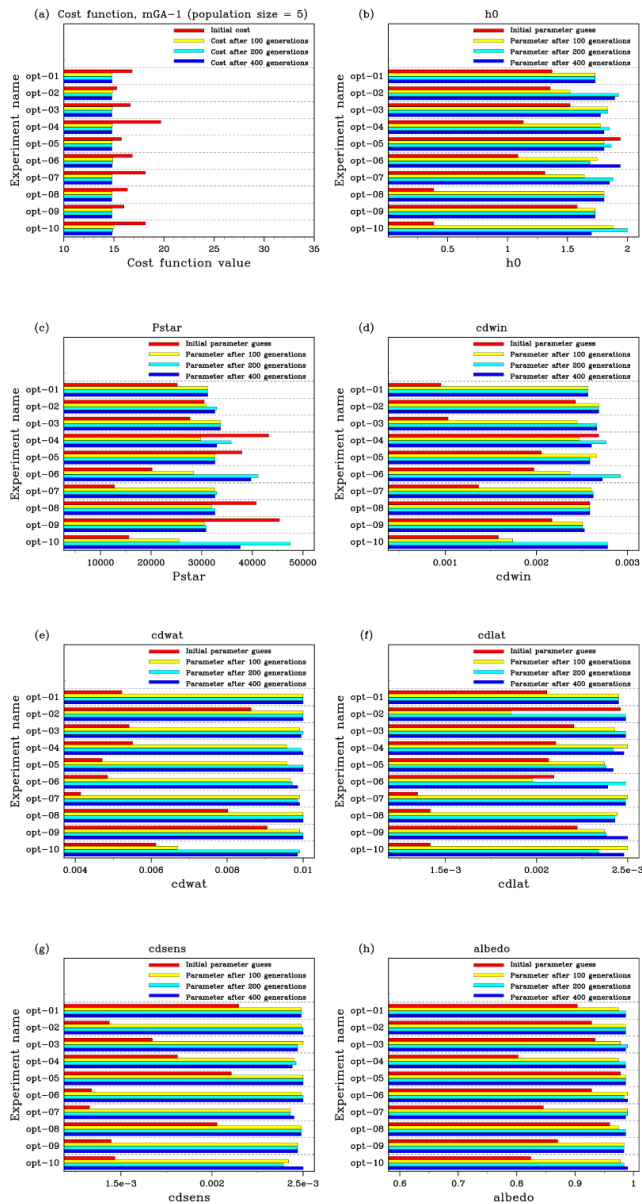
Fig. 14. The initial and optimized cost function values and associated parameter values obtained from 10 independent optimization experiments with the µGA (µGA – 1; population size = 5, number of possibilities = $(2^7)^7$). **(a)** Cost function values at the initial (red) and after 100 (yellow), 200 (green) and 400 (blue) generations. **(b–d)** Parameter values at the initial and after 100, 200 and 400 generations (the same color correspondence as **(a)**). The initial cost functions and associated parameter values for respective optimization experiments were obtained from the best fittest *individual* in the first generation.

The computational costs of the µGA limit the direct application to eddy-resolving models of the Arctic ($O \sim 1$ km). For this class of models, further development in high-performance-computing technology is needed.
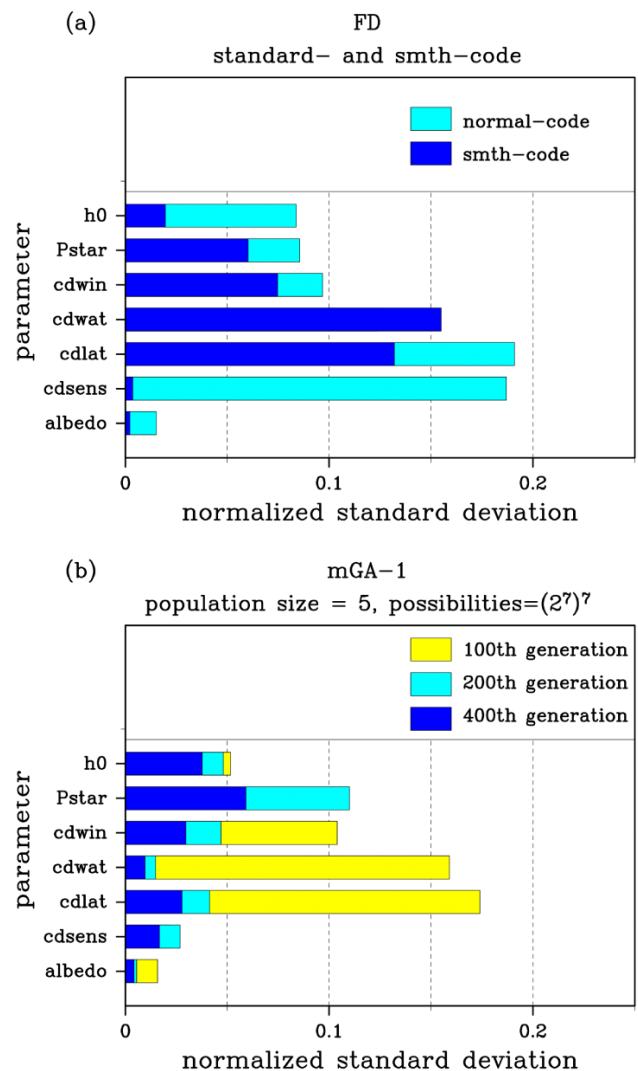

Fig. 15. Standard deviations of the estimated parameters obtained from 10 independent optimization experiments by **(a)** the FD method and **(b)** µGA. The standard deviations were normalized by corresponding parameter ranges: the normalized standard deviation of i-th parameter is given by $\left(m_i^{\max} - m_i^{\min}\right)^{-1} \left[L^{-1} \sum_{l=1}^{L} (\bar{m}_i - m_i)^2\right]^{\frac{1}{2}}$, where L is the number of optimization experiments and $\bar{m}_i$ is a mean of estimated i-th parameter.

Appendix A

An example of parameter renovation by the micro-genetic algorithm

Table A1 shows the parameters of 5 individuals and their fitness (cost function) for the first 6 generations of µGA – 1. In the first generation, all parameters are generated randomly and the fitness is evaluated by model runs. The fittest

Table A1. Example of evolution of estimated parameter values by the μGA (μGA−1).

				Generation 1				
No.	h_0	Pstar	cdwin	cdwat	cdlat	cdsens	alb	COST.FUNC.
1	1.16220	46456.69	0.000677	0.004142	0.001594	0.001427	0.67984	26.0288467
2	0.95276	41141.73	0.001327	0.008866	0.001772	0.001762	0.85488	17.9832013
3	1.13228	44330.71	0.002547	0.009811	0.002431	0.001427	0.79039	18.0514735
4	1.37165	25196.85	0.000953	0.005228	0.002057	0.002146	0.90402	16.8213995 ← the fittest
5	0.35433	44330.71	0.002272	0.007780	0.001476	0.002077	0.61228	26.8653972
				Generation 2				
No.	h_0	Pstar	cdwin	cdwat	cdlat	cdsens	alb	COST.FUNC.
1	1.37165	25196.85	0.000953	0.005228	0.002057	0.002146	0.90402	16.8213995 ← the old fittest
2	0.41417	23425.20	0.001661	0.006598	0.002352	0.001506	0.85488	17.1665073
3	1.37165	49291.34	0.002783	0.009764	0.002431	0.002146	0.70748	20.9840715
4	1.37165	19527.56	0.002331	0.008299	0.002037	0.002136	0.91630	15.8714103 ← the new fittest
5	1.31181	40787.40	0.000618	0.005606	0.002067	0.002096	0.85488	24.5029623
				Generation 3				
No.	h_0	Pstar	cdwin	cdwat	cdlat	cdsens	alb	COST.FUNC.
1	1.37165	19527.56	0.002331	0.008299	0.002037	0.002136	0.91630	15.8714103 ← the fittest
2	0.41417	26614.17	0.001701	0.009811	0.002352	0.001506	0.96543	17.1451719
3	0.41417	22362.20	0.000953	0.005276	0.002057	0.002136	0.90402	18.5657524
4	1.37165	25196.85	0.002370	0.005276	0.002037	0.002136	0.91630	17.0261603
5	1.37165	22362.20	0.001701	0.006409	0.002352	0.001506	0.85488	16.3909644
				Generation 4				
No.	h_0	Pstar	cdwin	cdwat	cdlat	cdsens	alb	COST.FUNC.
1	1.37165	22362.20	0.002370	0.004898	0.002352	0.002136	0.86717	18.2140693
2	1.37165	19527.56	0.002331	0.009811	0.002352	0.002136	0.91630	15.5498109 ← the new fittest
3	1.37165	19527.56	0.002331	0.008299	0.002037	0.002136	0.91630	15.8714103 ← the old fittest
4	0.41417	20944.88	0.001071	0.008299	0.002352	0.002136	0.96543	19.3435762
5	1.37165	25196.85	0.002370	0.005276	0.002037	0.002136	0.91630	17.0261603
				Generation 5				
No.	h_0	Pstar	cdwin	cdwat	cdlat	cdsens	alb	COST.FUNC.
1	1.37165	19527.56	0.002331	0.009811	0.002037	0.002136	0.91630	15.5990099
2	1.37165	19527.56	0.002331	0.008299	0.002352	0.002136	0.91630	15.8190923
3	1.37165	16692.91	0.002331	0.007921	0.002037	0.002136	0.86717	16.8763867
4	1.37165	19527.56	0.002331	0.009811	0.002352	0.002136	0.91630	15.5498109 ← the fittest
5	1.37165	25196.85	0.002331	0.004898	0.002352	0.002136	0.91630	17.2076177
				Generation 6				
No.	h_0	Pstar	cdwin	cdwat	cdlat	cdsens	alb	COST.FUNC.
1	1.37165	19527.56	0.002331	0.009811	0.002352	0.002136	0.91630	15.5498109 ← the fittest
2	1.47638	17047.24	0.001346	0.004236	0.001427	0.001496	0.71362	21.7795272 ← new individual
3	1.17717	28031.50	0.000657	0.004945	0.001378	0.001791	0.72898	23.4859782 ← new individual
4	0.98268	43622.05	0.002232	0.005276	0.001880	0.001614	0.75047	21.1250344 ← new individual
5	0.66850	43976.38	0.002508	0.005181	0.001644	0.001742	0.64606	25.7545548 ← new individual

individual (No. 4) is transferred to the second generation without any change (No. 1 of the second generation). In the second generation, a new fittest individual emerges by a random combination of the binary bits coming from the individuals of the first generation. Then the new fittest is transferred to the third generation, and the old fittest individual is abandoned. In the third generation, no new fittest is generated. It is evident in the third generation that the diversity of the parameter values is decreased. In the fifth generation, the difference of binary bits among the individuals is small and correspondingly the diversity of the parameters is also small.

Then reinitialization is done in the sixth generation producing new individuals.

Appendix B

Preliminary optimization experiment by the micro-genetic algorithm

In order to estimate the needed number of generations in relation to the number of possible parameter values (which is inversely proportional to the increment) and the expected accuracy of a solution, we performed optimization experiments with a pseudo-cost function.

The pseudo-function is defined as follows:

$$\text{cost} = \sum_{i=1}^{M} \left(\frac{m_i}{m_i^{\text{central value}}} - 1 \right)^2, \tag{B1}$$

where M is the number of optimized parameter ($M = 7$), m_i i-th parameter value and $m_i^{\text{central value}} = (m_i^{\text{max}} + m_i^{\text{min}})/2$ again the central value of the prescribed range of i-th parameter. The function is smooth and continuous everywhere and has only one minimum, cost $= 0$, at $\boldsymbol{m} = \boldsymbol{m}^{\text{central value}}$. A mean error of estimated parameters after a certain number of generations for a certain optimization trial is given by

$$e = \frac{1}{M} \sum_{m=1}^{M} \left(\frac{m_i^{\text{OPT}} - m_i^{\text{ANL}}}{m_i^{\text{max}} - m_i^{\text{min}}} \right), \tag{B2}$$

where e is a mean error normalized by the prescribed parameter ranges, m_i^{OPT} i-th parameter value after a certain optimization trial, and m_i^{ANL} the analytical solution of i-th parameter that gives the minimum cost function. m_i^{max} and m_i^{min} are the prescribed upper and lower limits of i-th parameter value given by Table 2. The expected value of the mean error after an optimization is given by

$$E = \sum_{j=1}^{J} e_j P\left(e_j\right), \tag{B3}$$

where E is the expectation of the mean error and $P(e_j)$ is a probability function.

Figure B1 shows the expected value of the mean error for different number of possibilities (proportional to reciprocal number of increment) and different number of generations. Each point in the figure (from 2^3 to 2^{15} possibilities for each number of generations) is calculated by 100 independent optimization trials ($J = 100$) by the μGA. In each optimization experiment, we can obtain e_j by Eq. (B2), while $P(e_j)$ is simply given by the reciprocal number of optimization experiments, $1/J$. The figure indicates that a larger number of generations generally gives a smaller expected error. On the other hand, the number of possibilities has a small contribution to the expected error in the range of possibilities larger
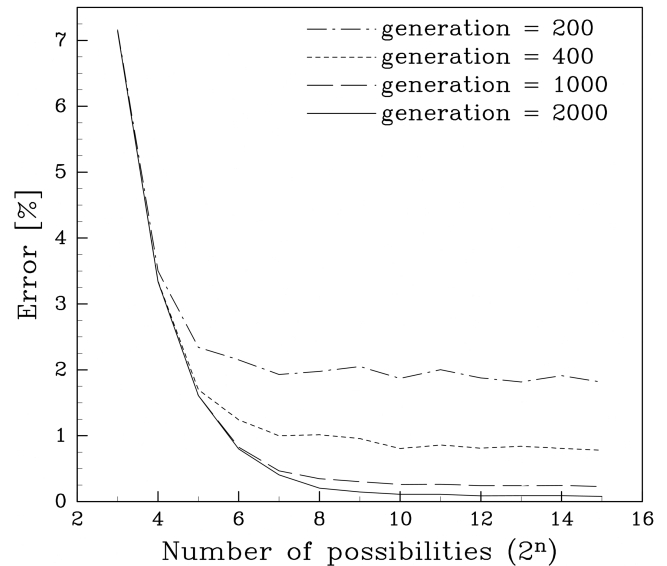

Fig. B1. An expected error of the estimated parameters by the μGA as a function of number of possibilities and number of generations, obtained from the idealized pseudo-function experiments. Each point in the figure is calculated by 100 independent optimization trials by the μGA. See text for details.

than 2^7. The figure also indicates that a solution with 1 % expected error can be obtained after 400 generations calculated with the number of possibilities of 2^7, while a solution with 0.5 % expected error needs 1000 generations with 2^7 possibilities. Note that these results are only valid for the simple cost function applied here, and it is not clear whether these results can be directly applicable to an optimization for a realistic cost function with complicated structure.

Appendix C

Evolution of estimated parameters by the micro-genetic algorithm

The evolution of the cost function and the corresponding parameters of 10 independent optimization experiments (μGA – 1) are shown in Fig. C1. The cost function and parameters are taken from the best fittest *individual* in the respective generations. The cost functions of all experiments rapidly reduce during the first 50 generations. Afterwards the reduction is moderate. However, the corresponding parameters still exhibit large variance in the 50th generation. A total of 200 generations are needed in some cases to reach the desired accuracy. After the 200th generation, the further reduction of the cost function is very small and also the parameters only change very slightly.
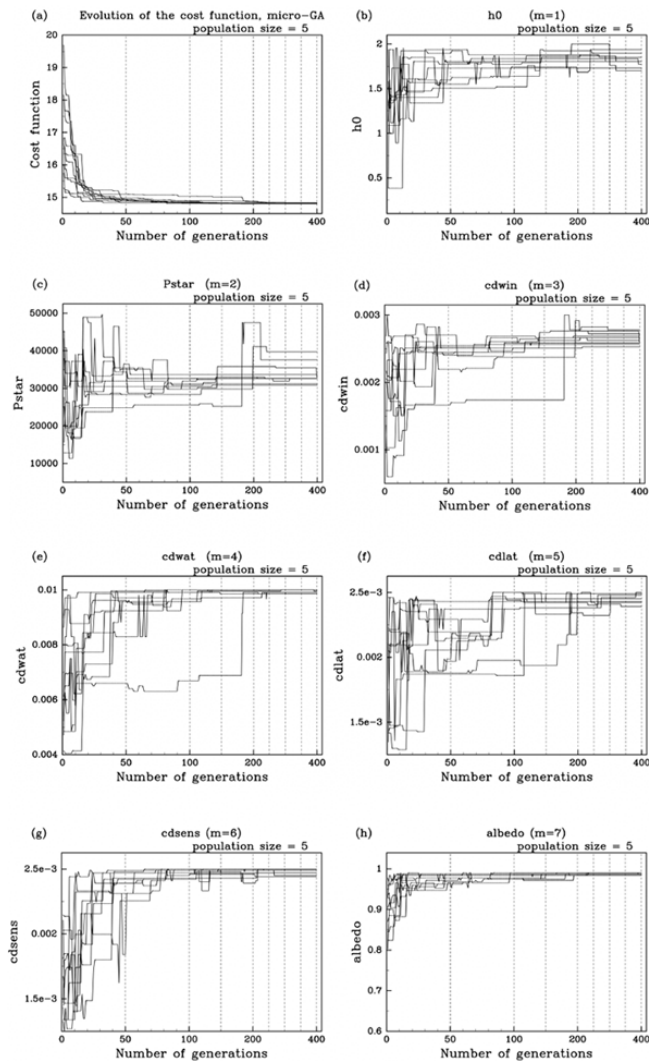
Fig. C1. Evolution of the cost function and associated parameter values through 10 independent optimization experiments by the μGA (μGA – 1: population size = 5, number of possibilities of each parameter = 2^7): **(a)** evolution of the cost function, **(b–h)** evolution of associated parameters **(b)** h_0, **(c)** P^*, **(d)** cdwin, **(e)** cdwat, **(f)** cdlat, **(g)** cdsens and **(h)** albedo. Note that the division of the horizontal axes is nonlinear. The vertical dashed lines were plotted every 50 generations.

References

Athias, V., Mazzega, P., and Jeandel, C.: Selecting a global optimization method to estimate the oceanic particle cycling rate constants, J. Mar. Res., 58, 675–707, 2000.

Bocquet, M.: Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var, Q. J. Roy. Meteorol. Soc., 138, 664–681, 2012.

Carroll, D. L.: Chemical Laser Modeling with Genetic Algorithms, AIAA J., 34, 338–346, 1996.

Cetin, B. C., Barhen, J., and Burdick, J. W.: Terminal Repeller Unconstrained Subenergy Tunneling (TRUST) for Fast Global Optimization, J. Optimiz. Theory App., 77, 97–126, 1993.

Chapman, W. L., Welch, W. J., Bowman, K. P., Sacks, J., and Walsh, J. E.:Arctic sea ice variability: Model sensitivities and a multi-decadal simulation, J. Geophys. Res., 99, 919–935, 1994.

Coello, C. A. C. and Pulido, G. T.: A Micro-Genetic Algorithm for multiobjective Optimization, in: EMO 2001, edited by: Zitzler, E., Thiele, L., Deb., K., Coello, C. A. C., and Corne, D., , LNCS 1993, Springer-Verlag, Berlin Heidelberg, 126–140, 2001.

Eastwood, S., Larsen, K. R., Lavergne, T., Nielsen, E., and Tonboe, R.: Global Sea Ice Concentration Reprocessing Product User Manual, Product OSI-409, The EUMETSAT Network of Satellite Application Facilities, 2010.

Eisenman, I., Untersteiner, N., and Wettlaufer, J. S.: On the reliability of simulated Arctic sea ice in global climate models, Ge ophys. Res. Lett., 34, L10501, doi:10.1029/2007GL029914, 2007.

Evensen, G. and Fario, N.: Solving for the Generalized Inverse of the Lorenz Model, Journal of Meteorological Society of Japan, 75, 229–243, 1997.

Garcia-Gorriz, E., Hoepffner, N., and Ouberdous, M.: Assimilation of SeaWiFS data in a coupled physical – biological model of the Adriatic Sea, J. Marine Sys., 40–41, 233–252, doi:10.1016/S0924-7963(03)00020-4, 2003.

Gerdes, R. and Köberle, C.: Comparison of Arctic sea ice thickness variability in IPCC Climate of the 20th Century experiments and in ocean-sea ice hindcasts, J. Geophys. Res., 112, C04S13, doi:10.1029/2006JC003616, 2007.

Gerdes, R., Karcher, M. J., Kauker, F., and Schauer, U.: Causes and development of repeated Arctic Ocean warming events, Geophys. Res. Lett., 30, 1980, doi:10.1029/2003GL018080, 2003.

Goldberg, D. E.: Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA, 432 pp., 1989.

Harder, M.: Rauhigkeit und Alter des Meereises in der Arktis–Numerische Untersuchungen mit einem grossskaligen Modell, Ph.D. Thesis, Berichte zur Polarforschung, 203, Alfred-Wegener-Institut, 1996 (in German with English summary).

Harder, M. and Fischer, H.: Sea ice dynamics in the Weddell Sea simulated with an optimized model, J. Geophys. Res., 104, 11151–11162, 1999.

Hibler III, W. D.: A Dynamic Thermodynamic Sea Ice Model, J. Phys. Oceanogr., 9, 815–846, 1979.

Hibler III, W. D. and Bryan, K.: A diagnostic ice-ocean model, J. Phys. Oceanogr., 17, 987–1015, 1987.

Holland, D. M., Mysak, L. A., Manak, D. K., and Oberhuber, J. M.: Sensitivity Study of a Dynamic Thermodynamic Sea Ice Model, J. Geophys. Res., 98, 2561–2586, 1993.

Holland, J. H.: Adaption in Natural and Artificial System, A Bradfford Book, Univ. Michigan Press, 1975.

Hunke, E. C. and Dukowicz, J. K.: An Elastic-Viscous-Plastic Model for Sea Ice Dynamics, J. Phys. Oceanogr., 27, 1849–1867, 1997.

Hunke, E. C. and Lipscomb, W. H.: CICE: the Los Alamos sea ice model, documentation and software user's manual, LACC-98-16 v. 3. Los Alamos National Laboratory, NM, 2001.

Huret, M., Gohin, F., Delmas, D., Lunven, M., and Garçon, V.: Use of SeaWiFS data for light availability and parameter estimation of a phytoplankton production model of the Bay of Biscay, J. Marine Syst., 65, 509–531, 2007.

Johnson, E. G. and Abushagur, M. A. G.: Microgenetic-algorithm optimization methods applied to dielectric gratings, J. Opt. Soc. Am. A 12, 1152–1160, 1995.

Johnson, M., Gaffigan, S., Hunke, E., and Gerdes, R.: A comparison of Arctic Ocean sea ice concentration among the coordinated AOMIP model experiments, J. Geopys. Res., 112, C04S11, doi:10.1029/2006JC003690, 2007.

Johnson, M., Proshutinsky, A., Aksenov, Y., Nguyen, A. T., Lindsay, R., Hass, C., Zhang, J., Diansky, N., Kwok, R., Maslowski, W., Häkkinen, S., Ashik, I., and de Cuevas, B.: Evaluation of Arctic sea ice thickness simulated by Arctic Ocean Model Intercomparison Project models, J. Geophys. Res., 117, C00D13, doi:10.1029/2011JC007257, 2012.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471, 1996.

Karcher, M. J., Gerdes, R., Kauker, F., and Köberle, C.: Arctic warming: Evolution and spreading of the 1990s warm event in the Nordic seas and the Arctic Ocean, J. Geophys. Res., 108, 3034, doi:10.1029/2001JC001265, 2003.

Karcher, M., Gerdes, R., Kauker, F., Köberle, C., and Yashayaev, I.: Arctic Ocean change heralds North Atlantic freshening, Geophys. Res. Lett., 32, L21606, doi:10.1029/2005GL023861, 2005.

Karcher, M., Kauker, F., Gerdes, R., Hunke, E., and Zhang, J.: On the dynamics of Arctic Water circulation in the Arctic Ocean, J. Geophys. Res., 112, C04S02, doi:10.1029/2006JC003630, 2007.

Karcher, M., Beszczynska-Möller, A., Kauker, F., Gerdes, R., Heyen, S., Rudels, B., and Schauer, U.: Arctic Ocean warming and its consequences for the Denmark Strait overflow, J. Geophys. Res., 116, C02037, doi:10.1029/2010JC006265, 2011.

Kauker, F., Gerdes, R., Karcher, M., Köberle, C., and Lieser, J. L.: Variability of Arctic and north Atlantic sea ice: A combined analysis of model results and observations from 1978 to 2001, J. Geophys. Res., 108, 3182, doi:10.1029/2002JC001573, 2003.

Kauker, F., Gerdes, R., Karcher, M., and Köberle, C.: Impact of North Atlantic Current changes on the Nordic Seas and the Arctic Ocean, J. Geophys. Res., 110, C12002,

doi:10.1029/2004JC002624, 2005.

Kauker, F., Kaminski, T., Karcher, M., Giering, R., Gerdes, R., and Voßbeck, M.: Adjoint analysis of the 2007 all time Arctic sea-ice minimum, Geophys. Res. Lett., 36, L03707, doi:10.1029/2008GL036323, 2009.

Kim, J. G., Hunke, E. C., and Lipscomb, W. H.: Sensitivity analysis and parameter tuning scheme for global sea-ice modeling, Ocean Model., 14, 61–80, 2006.

Kim, Y., Gotoh, K., Toyosada, M., and Park, J.: Micro-Genetic Algorithms (μGAs) for Hard Combinatorial Optimization Problems, Proceedings of the Twelfth International Offshore and Polar Engineering Conference, 230–235, 2002.

Kirkpatrick, S., Gelatt Jr., C. D., and Vecchi, M. P.: Optimization by Simulated Annealing, Science, 220, 671–680, 1983.

Köberle, C. and Gerdes, R.: Mechanisms Determining the Variability of Arctic Sea Ice Conditions and Export, J. Climate, 16, 2843–2858, 2003.

Krishnakumar, K.: Micro-genetic algorithms for stationary and non-stationary function optimization, Intel. Control Adaptive Syst., 1996, 289–296, 1989.

Kwok, R. and Cunningham, G. F.: ICESat over Arctic sea ice: Estimation of snow depth and ice thickness, J. Geophys. Res., 113, C08010, doi:10.1029/2008JC004753, 2008.

Kwok, R., Cunningham, G. F., Zwally, H. J., and Yi, D.: Ice, Cloud, and land Elevation Satellite (ICESat) over Arctic sea ice: Retrieval of freeboard, J. Geophys. Res., 112, C12013, doi:10.1029/2006JC003978, 2007.

Kwok, R., Cunningham, G. F., Wensnahan, M., Rigor, I., Zwally, H. J., and Yi, D.: Thinning and volume loss of the Arctic Ocean sea ice cover: 2003–2008, J. Geophys. Res., 114, C07005, doi:10.1029/2009JC005312, 2009.

Lavergne, T. and Eastwood, S.: Low Resolution Sea Ice Drift Product User's Manual, GBL LR SID – OSI 405, The EUMETSAT Network of Satellite Application Facilities, 2010.

Ledley, T. S.: The Climate Response to Meridional Sea-Ice Transport, J. Climate, 4, 147–163, 1991a.

Ledley, T. S.: Snow on Sea Ice: Competing Effects in Shaping Climate, J. Geophys. Res., 96, 17195–17208, 1991b.

Lipscomb, W. H., Hunke, E. C., Maslowski, W., and Jakacki, J.: Ridging, strength, and stability in high-resolution sea ice models, J. Geophys. Res., 112, C03S91, doi:10.1029/2005JC003355, 2007.

Liu, D. C. and Nocedal, J.: On the limited memory BFGS method for large scale optimization, Math. Program., 45, 503–528, 1989.

Lu, P., Li, Z., Cheng, B., and Leppäranta, M.: A parameterization of the ice-ocean drag coefficient, J. Geophys. Res., 116, C07019, doi:10.1029/2010JC006878, 2011.

Lüpkes, C. and Birnbaum, G.: Surface drag in the arctic marginal sea-ice zone: a comparison of different parameterization concepts, Bound.-Lay. Meteorol., 117, 179–211, doi:10.1007/s10546-005-1445-8, 2005.

Lv, W., Feng, E., and Lei, R.: Parameter identification for a nonlinear non-smooth thermodynamic system of sea ice, Int. J. Therm. Sci., 48, 195–203, 2009.

Martin, T. and Gerdes, R.: Sea ice drift variability in Arctic Ocean Model Intercomparison Project models and observations, J. Geophys. Res., 112, C04S10, doi:10.1029/2006JC003617, 2007.

Mazzega, P.: On the assimilation and inversion of small data sets under chaotic regimes, in: Inverse Methods in Global Biogeo-

chemical Cycles, edited by: Kasibhatla, P., Heimann, M., Rayner, P., Mahowald, N., Prinn, R. G., and Hartley, D. E., Geophys. Monogr. Ser., 114, 223–237. doi:10.1029/GM114p0223, 2000.

McPhee, M.: Air–Ice Ocean Interaction: Turbulent Ocean Boundary Layer Exchange Process, Springer, 266 pp., 2008.

Menemenlis, D., Fukumori, I., and Lee, T.: Using Green's Functions to Calibrate an Ocean General Circulation Model, Mon. Weather Rev., 133, 1224–1240, 2005.

Miller, P. A., Laxon, S. W., Feltham, D. L., and Gresswell, D. J.: Optimization of a Sea Ice Model Using Basinwide Observations of Arctic Sea Ice Thickness, Extent, and Velocity, J. Climate, 19, 1089–1108, 2006.

Mochizuki, T., Igarashi, H., Sugiura, N., Masuda, S., Ishida, N., and Awaji, T.: Improved coupled GCM climatologies for summer monsoon onset studies over Southeast Asia, Geophys. Res. Lett., 34, L01706, doi:10.1029/2006GL027861, 2007.

Nguyen, A. T., Menemenlis, D., and Kwok, R.: Arctic ice-ocean simulation with optimized model parameters: Approach and assessment, J. Geophys. Res., 116, C04025, doi:10.1029/2010JC006573, 2011.

Pacanowski, R. C.: MOM 2 documentation, user's guide and reference manual, GFDL Ocean Group Tech. Rep. 3, Geophysical Fluid Dynamics Laboratory, Princeton Univ., Princeton, NJ, 1995.

Rothrock, D. A., Zhang, J., and Yu, Y.: The arctic ice thickness anomaly of the 1990s: A cosistent view from observations and models, J. Geophys. Res., 108, 3083, doi:10.1029/2001JC001208, 2003.

Schartau, M. and Oschlies, A.: Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I – Method and parameter estimates, J. Mar. Res., 61, 765–793, 2003.

Shigemitsu, M., Okunishi, T., Nishioka, J., Sumata, H., Hashioka, T., Aita, M. N., Smith, S. L., Yoshie, N., Okada, N., and Yamanaka, Y.: Development of a one-dimensional ecosystem model including the iron cycle applied to the Oyashio region, western subarctic Pacific, J. Geophys. Res., 117, C06021, doi:10.1029/2011JC007689, 2012.

Shine, K. P. and Henderson-Sellers, A.: The Sensitivity of a Thermodynamic Sea Ice Model to Changes in Surface Albedo Parameterization, J. Geophys. Res., 90, 2243–2250, 1985.

Steele, M., Morley, R., and Ermold, W.: PHC: A global ocean hydrography with a high-quality Arctic Ocean, J. Climate, 14, 2079–2087, 2001.

Stevens, D. P.: The open boundary condition in the United Kingdom Fine-Resolution Antarctic Model, J. Phys. Oceanogr., 21, 1494–1499, 1991.

Thomas, D. N. and Dieckmann, G. S.: Sea Ice, 2nd Edn., Wiley-Blackwell, 640 pp., 2009.

Vallino, J. J.: Improving marine ecosystem models: Use of data assimilation and mesocosm experiments, J. Mar. Res., 58, 117–164, 2000.

Wadhams, P.: Ice in the Ocean, RoutledgeCurzon, 364 pp., 2000.

Zhang, J. and Hibler III, W. D.: On and efficient numerical method for modeling sea ice dynamics, J. Geophys. Res., 102, 8691–8702, 1997.