LA-UR-15-29485 (Accepted Manuscript)

Uncertainty quantification and global sensitivity analysis of the Los Alamos sea ice model

Urrego Blanco, Jorge Rolando
Urban, Nathan Mark
Hunke, Elizabeth Clare
Turner, Adrian Keith
Jeffery, Nicole

To be published in: Journal of Geophysical Research: Oceans

DOI to publisher's version: 10.1002/2015JC011558

Permalink to record: http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-15-29485

Research Article
Journal of Geophysical Research: Oceans
DOI 10.1002/2015JC011558

Uncertainty Quantification and Global Sensitivity Analysis of the Los Alamos Sea Ice Model

Jorge R. Urrego-Blanco, Nathan M. Urban, Elizabeth C. Hunke, Adrian K. Turner, and Nicole Jeffery

Los Alamos National Laboratory, Los Alamos, New Mexico, 87545, USA

Abstract. Changes in the high latitude climate system have the potential to affect global climate through feedbacks with the atmosphere and connections with mid latitudes. Sea ice and climate models used to understand these changes have uncertainties that need to be characterized and quantified. We present a quantitative way to assess uncertainty in complex computer models, which is a new approach in the analysis of sea ice models. We characterize parametric uncertainty in the Los Alamos sea ice model (CICE) in a standalone configuration and quantify the sensitivity of sea ice area, extent and volume with respect to uncertainty in 39 individual model parameters. Unlike common sensitivity analyses conducted in previous studies where parameters are varied one at a time, this study uses a global variance-based approach in which Sobol' sequences are used to efficiently sample the full 39-dimensional parameter space. We implement a fast emulator of the sea ice model whose predictions of sea ice extent, area, and volume are used to compute the Sobol' sensitivity indices of the 39 parameters. Main effects and interactions among the most influential parameters are also estimated by a non-parametric regression technique based on generalized additive models. A ranking based on the sensitivity indices indicates that model predictions are most sensitive to snow parameters such as conductivity and grain size, and the drainage of melt ponds. It is recommended that research be prioritized towards more accurately determining these most influential parameter values by observational studies or by improving parameterizations in the sea ice model.

Received: Dec 14, 2015; Revised: Feb 23, 2016; Accepted: Mar 21, 2016

1. Introduction

High latitudes are important components of the climate system where changes occur more rapidly than at lower latitudes and can have significant effects on the global climate [*Screen and Simmonds*, 2010]. Such effects can be amplified by feedbacks between the ocean-ice-atmosphere system and through connections with mid and lower latitudes [*Vage et al.,* 2008; *Francis et al.*, 2009]. To understand these processes, Earth system models are used which typically include atmosphere, ocean, ice and land components. Sea ice models are not only important in the context of climate dynamics but also for operational forecasts at high latitudes [*Blockley et al.*, 2014; *Dupont et al.*, 2015]. . Sea ice models have different levels of complexity that typically include a significant number of physics-based processes. Sea ice model estimates, like any other models, are expected to deviate to some degree from the exact representation of the real world because of uncertainties associated with model predictions. It is important to establish and understand potential sources of uncertainty and flaws in the models used in climate research and operational forecast systems, so that conclusions drawn in climate studies are robust and transparent, and operational predictions adequate. This should improve, for instance, seasonal ice forecasts and our ability to project future impacts of sea ice loss.

Uncertainty in models can arise from multiple sources [*Kennedy and O'Hagan*, 2001], including parametric uncertainty, which is caused by the lack of knowledge of the exact values that model parameters should take in the simulations. For sea ice models, parameterizations of physical processes include a number of parameters for which accurate values are not always well established. The Los Alamos sea ice model (CICE 5.1), for instance, has implemented new or complex parameterizations, which include a delta-Eddington radiation scheme [*Holland et al.*, 2012], a new melt pond formulation [*Hunke et al.*, 2013], a mushy thermodynamic model [*Turner et al.*, 2013], and a variable drag coefficient scheme [*Tsamados et al.*, 2014]. There is a need to examine the sensitivity of model predictions to uncertainties in input parameters. A comprehensive sensitivity analysis should help to identify important physical processes affecting sea ice distributions, and determine what model components explain most of the model uncertainty.

It has been common practice in climate studies to assess the sensitivity of models to reduced sets of input parameters, or only over limited ranges of parameter variation. The most common practice is to conduct one-at-a-time (OAT) sensitivity analyses in which one parameter is varied with respect to a control experiment, while others are kept constant, usually at default values. This approach, however, cannot identify interactions among parameters and assumes linearity and additivity in climate models, which are strong limitations given the complexity of the climate system. The common approach only examines the main effects of parameters, and does not quantify model uncertainty due to individual parameters. An objective ranking of important parameters is therefore not possible in OAT analysis. For sea ice models, in particular, *Peterson et al*. [2010] examined sensitivity of sea ice to 10 model parameters in two standalone sea ice model configurations, using a linear regression approach and the standardized regression coefficients as measurements of model sensitivity. *Uotila et al.* [2012] also examined the sensitivity of sea ice in a coupled ocean-ice model to 10 parameters using 100 model runs. They explored the parameters sampling at discrete levels within prescribed ranges of variation, and included a reduced subset of model parameters used in the delta-Eddington radiation scheme. A sensitivity study using a global coupled climate model [*Rae et al.*, 2014], explored sea ice sensitivity to 16 parameters, examining departures of 16 experiments from model results in a control run simulation. Other sensitivity analyses applied to sea ice models have been based on automatic differentiation [*Kim et al.*, 2006] or the use of adjoint models. These methods, however, are mainly local, which means the sensitivity is examined only in the vicinity of certain parameter configurations without fully exploring the entire parameter space in the model, which can have serious limitations on high dimensional parameter spaces [*Saltelli and Annoni*, 2010]. A full exploration of high-dimensional parameter spaces in climate models is prohibitive in conventional sensitivity analyses because of the computing resources required to conduct a single model run. An optimal approach, called global sensitivity analysis (GSA), is one in which all input parameters are taken into account, examining possible interactions among them, and exploring parameters over all the plausible values they can take.

In this study we use GSA to examine the sensitivity of the CICE sea ice model in standalone mode to the parameters contributing to uncertainty in model predictions. It could be expected that sensitivities in a standalone configuration differ to some extent from those in a coupled

climate system. The approach in this study, however, has the advantage of requiring significantly less computational resources while still being able to elucidate important physical processes affecting sea ice. The approach uses quasi-random sequences to sample the parameter space, and makes use of a fast surrogate of the sea ice model to acquire the large amount of data necessary to estimate sensitivity. This approach uses a large ensemble of model runs, allowing us to conduct the sensitivity analysis at an unprecedented level of detail. Our objectives are to determine the most important model parameters contributing to uncertainty in predictions of sea ice, emphasizing snow and radiation parameters, and other new model developments since *Kim et al. [2006]*, and to elucidate underlying physical processes of these parameters in terms of their individual effects as well as the most important interactions among them. Section 2 presents the GSA methodology used in this study, which is based on the variance decomposition of model predictions. The methodology used to characterize model uncertainty and quantify the model sensitivity is presented in Section 3. In Section 4 we present the results of the uncertainty propagation through the model onto model predictions, the implementation of a surrogate of the sea ice model, and estimates of sensitivity indices. A discussion of the sensitivity indices, and their implications, is presented in Section 5, with conclusions in Section 6.

2. Variance-based Global Sensitivity Analysis

The purpose of a Global Sensitivity Analysis (GSA) is to determine the contribution of uncertainty in individual model parameters to the model uncertainty, allowing individual model parameters to vary over the entire range of plausible values they may take. The uncertainty of a model parameter can be characterized by a probability distribution (section 3.2). The term "global" in this context implies that an efficient way of sampling must be adopted so that uncertainties in all parameters are sampled jointly. Importantly, no assumptions are made about the linearity or additivity of the model, implicit in most commonly used sensitivity analyses.

The variance of model predictions over input parameter distributions is used as a measurement of sensitivity. Given a model prediction, a variance-based GSA explicitly seeks to apportion its variance to individual parameters. Given p model input parameters, a p-dimensional parametric space can be uniformly distributed and normalized in a unit hypercube

$$\Omega^p = (\mathbf{X}|0 < X_i < 1; \quad i = 1, \ldots, p)$$

A Hoeffding decomposition [*Saltelli et al*., 2010] of a model prediction $Y = f(\mathbf{X})$ can be expressed as a sum of 2^p terms of increasing dimensionality [*Sobol'*, 2001]

$$f(\mathbf{X}) = f_0 + \sum_{i=1}^{p} f_i(X_i) + \sum_{1 \le i < j}^{p} f_{ij}(X_i, X_j) + \cdots + f_{1,2,\ldots,p}(X_1, X_2, \ldots, X_p) \quad (1)$$

where the integrals of the functions in the decomposition over any of their variables must be zero, and therefore the functions be mutually orthogonal [*Jacques et al., 2006*]. It follows that variances can also be decomposed according to

$$V(Y) = \sum_{i=1}^{p} V_i + \sum_{1 \le i < j}^{p} V_{ij} + \cdots + V_{1,2,\ldots,p} \quad (2)$$

Normalizing the above equation by the total variance *V(Y),* one obtains

$$\sum_{i=1}^{p} S_i + \sum_{1 \le i < j}^{p} S_{ij} + \cdots + S_{1,2,\ldots,p} = 1 \quad (3)$$

where each term represents a fraction of the total variance and thus can be used to estimate the sensitivity of model predictions to input parameters. The terms in the first sum are the first-order sensitivity indices, the terms in the second sum are the second-order sensitivity indices, and so on [*Homma and Saltelli*, 1996]. The first-order indices, also called main effects, indicate the reduction in *V(Y)* that would be obtained if parameter X_i could be set to a fixed value. The main effects are useful to prioritize parameters for uncertainty reduction, thus guiding research toward determining more accurate values for parameters with large first-order sensitivity indices.

The higher-order indices indicate variance explained by interactions between parameters, which cannot be attributed to individual effects of input parameters. For instance, S_{ij} accounts for the

variance explained by the interaction between X_i and X_j so that a non-zero S_{ij} indicates that the effect of X_i on Y would also depend on what value X_j takes.

A useful measurement of sensitivity is the total sensitivity index (S_{Ti}) which is the sum of all the terms involving x_i in equation (3). S_{Ti} can also be thought of as the variance that would remain if all input parameters but x_i could be set to fixed values. The total index is useful to determine which input parameters are not very important (either directly or through their interactions) for the model uncertainty, so that parameters with small total indices can be fixed at default values. S_i and S_{Ti} have also been expressed as [*Saltelli et al., 2010*]

$$S_i = \frac{V_{\mathbf{X}_i}(E_{\mathbf{x}_{\sim i}}(Y|X_i))}{V(Y)} \quad (4)$$

and

$$S_{Ti} = \frac{E_{\mathbf{X}_{\sim i}}(V_{\mathbf{X}_i}(Y|X_{\sim i}))}{V(Y)} = 1 - \frac{V_{\mathbf{X}_{\sim i}}(E_{\mathbf{X}_i}(Y|X_{\sim i}))}{V(Y)} \quad (5)$$

where X_i represents the i^{th} input parameter, and $V_{\mathbf{X}_i}$ and $E_{\mathbf{X}_i}$ indicate variances and expectations taken over all values of X_i. Similarly $X_{\sim i}$ represents all input parameters but X_i, and $V_{\mathbf{x}_{\sim i}}$ and $E_{\mathbf{x}_{i\sim}}$ indicate variances and expectations taken over all parameters keeping X_i fixed.

In practice, estimation of sensitivity indices given by equations 4 and 5 requires a large number of model evaluations. The method for estimating the variances and sensitivity indices in these equations is presented in the Appendix.

3. Methods

We use CICE in standalone mode (section 3.1) to examine sea ice sensitivity to a much larger number of input parameters than has been addressed in previous studies. The model output variables of interest are sea ice extent, area and volume integrated over the northern and southern hemispheres. Area refers to actual area of ocean covered by sea ice, while extent refers to the area covered by more than 15% sea ice. We initially identified 49 input parameters associated with the CICE configuration described in section 3.1, whose values are not accurately known and are therefore uncertain. We chose 39 of these input parameters (Table 1) to conduct the GSA, in order to examine as many uncertain parameters as possible, while neglecting 10 parameters from the analysis because they were specific to the ocean, or because there was enough confidence that they would turn out to be negligibly important for the variables of interest. The calculation of sensitivity indices for the 39 parameters left in the analysis would require a very large number of model evaluations and a full exploration of the parametric space over the entire ranges of plausible parameter values (section 3.3). In practice, this is computationally prohibitive for the sea ice model configuration used in this study and therefore the use of a statistical emulator is explored to produce data points at a reduced computational cost (section 3.5). Next we introduce the main features of CICE, the criteria used to design the ensemble of different model configurations, the implementation of a Gaussian process emulator to approximate the CICE model, and the methodology used to estimate the sensitivity indices.

3.1. Los Alamos Sea Ice Model

CICE is one of the most widely used sea ice models for climate research [*Gent et al.*, 2011; *Hewitt et al*., 2011; *Bentsen et al.*, 2013] and operational forecast systems at high latitudes [*Tonani et al.,* 2015]. The model solves dynamic and thermodynamic equations for multiple ice thickness categories [*Hunke et al.*, 2015]. For this analysis, in the dynamic component an elastic-viscous-plastic (EVP) rheology is used to account for deformation of the ice pack [*Hunke and Dukowicz,* 1997], and sea ice is redistributed among thickness categories due to mechanical redistribution [*Thorndike et al., 1975; Liscomb et al.*, 2007]. The thermodynamic component describes the ice pack as a mushy layer [*Turner et al.*, 2013] in which desalination occurs as sea ice grows; temperature and salinity are prognostic variables over seven ice layers on each of five

thickness categories used in this study. Albedo and radiative fluxes at the surface of sea ice are computed using a delta-Eddington radiation scheme based on inherent optical properties defining scattering and absorption properties of the snow-ice surface [*Briegleb and Light*, 2007; *Holland et al.*, 2012]. In the present model configuration, melt water and liquid precipitation falling over sea ice are tracked using a level-ice pond formulation in which melt ponds are carried as tracers on undeformed sea ice, and a depth-area ratio is assumed for changes in pond volume [*Hunke et al.*, 2013].

In this study we use CICE in standalone mode with a horizontal resolution of 1º [*Hunke and Holland,* 2007]. The model's slab ocean mixed layer is forced by a monthly climatological ocean forcing derived from a Community Climate System Model (CCSM) run [*Collins, et al.*, 2006]. The oceanic forcing fields consist of sea surface slope and currents, salinity and deep ocean heat fluxes. Atmospheric forcing fields of 6-hourly wind speed, air temperature, and specific humidity are taken from the coordinated ocean-ice reference experiment (CORE 2) [*Griffies et al.*, 2009; *Large and Yeager*, 2009]. The atmospheric forcing also includes the specification of monthly climatologies of precipitation [*Griffies et al.*, 2009] and cloud fraction [*Röske, 2001*]. The model is integrated from 1958-1975, with no ice as initial condition. In 1975 several model runs with different configurations (described in the experimental design of section 3.3) are restarted and run until 2009. Only model results from 1980-2009 are used in the sensitivity analysis to allow the model to adjust after the 1975 restart.

3.2. Prior distributions of uncertain parameters in CICE

The CICE configuration described above has a number of uncertain parameters that were broadly categorized in terms of the physical processes in which they are involved (Table 1). Uncertainties associated with 39 parameters examined in this study are characterized through probability distributions, including the minimum and maximum values that individual parameters can take. In order to assign prior probability distributions to the parameters (representing our prior knowledge about them), we determined plausible values that these parameters might take based on a literature review of experimental data, model studies, and expert judgment.

Probability distributions for the model parameters are not generally available, except for a few parameters such as ice and snow density. Different types of probability distributions are used to characterize uncertainty in the input parameters. Uniform distributions are used when there is a considerably lack of information about a parameter's actual distribution. Normal distributions are assigned if standard deviation and mean values are available from the literature review, and observations support Gaussian behavior. In cases when not much information is available for a given parameter, but different references suggest unimodal distributions centered at a preferential value, a triangular distribution is chosen. Whenever a parameter is constrained to be positive, or varies across several orders of magnitude, log-uniform, log-normal, or log-triangular distributions are considered. For parameters representing ratios, ranging between zero and one, logit-normal distributions are considered. Where there is enough information, we center the distributions at the mean values found in the different studies, and the parameter's minimum and maximum values are set at ± 3 standard deviations from that mean. If these extremes lie far away from realistic values, or do not appear to be plausible, we set them in a more ad-hoc manner. A summary of each parameter's prior (pdf, minimum, and maximum values) specified in this study is presented in Table 2.

The distributions describing uncertainty in input parameters (Table 2) are used to design combinations of parameters to be used in the GSA. In particular, a 39-dimensional unit hypercube is used in the next section to produce ensemble designs. First, the coordinate of a given parameter in a unit hypercube is mapped onto an ordinate in the closed interval defined by the cumulative distribution functions of the minimum and maximum values for that parameter. Second, the corresponding parameter value is obtained by taking the inverse cumulative distribution function of the ordinate previously obtained. More details about the methods used to design the sets of parameter combinations used in this study are described in the following section.

3.3. Experimental design

In GSA the full exploration of high dimensional parametric spaces requires the sampling of data points to be done efficiently to keep the number of model evaluations as small as possible. It is a common practice to use low-discrepancy sequences as a type of quasi Monte Carlo sampling in the design of sampling matrices (sequences of data points consisting of different combinations of input parameter values). Low-discrepancy sequences are constructed to distribute data points uniformly over the parametric spaces, and include Latin Hypercubes, Niederreiter, Halton, and Sobol' sequences [*Lemieux*, 2009]. In this study we have used scrambled Sobol' sequences [*Owen*, 1998] which allow good pairwise distributions in high dimensions with the added feature that additional data points added to the sequence guarantee higher density sampling of the parameter space.

Using scrambled Sobol' sequences, model ensembles are created for 39 dimensions, the same number of input parameters. In each ensemble member, all of the 39 input parameters change their values so that a large number of sampling points in parameter space is possible. We first use the Sobol' design, along with the prior distributions in Table 2, to conduct 150 CICE model runs and make a preliminary screening of input variables -output variables relationships (section 3.4). Then 400 Sobol' points are added to the initial sequence, adding up to a total of 550 model runs, whose results are used to train a surrogate of the model. The first 400 members of this extended Sobol' sequence (including those in the preliminary screening) are transformed onto parameter space using the priors in Table 2. The last 150 members are transformed using uniform distributions for all of the 39 parameters, to increase the sampling at the tails of the prior distributions (section 3.5). Finally, we use Sobol' sequences to produce large sampling matrices ($\sim 10^5$ members) for use in the estimation of the sensitivity indices (section 4.3).

3.4. Preliminary screening

The calculation of sensitivity indices requires a very large number of model evaluations. Test cases conducted on highly non-linear analytical functions with $O(10)$ parameters have shown that the number of model evaluations required to obtain the total sensitivity indices is $O(10^4)$ [*Saltelli et al.*, 2010]. This number of model evaluations would be computationally prohibitive for a complex and computationally intensive model such as CICE, which depends on four times

that many input parameters. We seek an emulation to accomplish fast estimates of model predictions (section 3.5). However, emulators of high dimensional functions might not scale well and can be very hard to construct [*Bengio et al.*, 2006]. If the subset of input parameters that accounts for global variations of a function (or active variables) is large, then building an emulator can be impractical [*O'Hagan and West*, 2013].

As was mentioned in section 3.3, we initially conducted a set of 150 CICE model evaluations to preliminarily screen the model predictions, assess the suitability of implementing a statistical emulator for sea ice extent, area, and volume in CICE, and to identify inactive parameters that could be fixed in the analysis to alleviate the computational cost. In this approach, we look for evidence of non-random relationships between the model predictions and each of the 39 input parameters using a combined squared rank differences and Spearman's rank correlation test [*Hora and Helton*, 2003]. The method is useful to detect monotone relations as well as general non-linear patterns in datasets of model inputs and model predictions.

Of the initial 150 runs, about 100 were successful and could be used for the preliminary screening. Other runs failed due to parameter-induced instabilities. From the screening, the number of active variables (for which non-random behavior was detected) per output variable of interest ranges somewhere between 4 and 12 parameters (not shown), which indicates that building an emulator might be feasible for these types of variables. The preliminary screening also showed that a relatively large number of parameters displayed non-random relationships with variables of interest at some point throughout the year, therefore supporting our choice of not fixing any of the 39 parameters made in the experimental design.

3.5. Gaussian Process Emulator

The predictions from the CICE model are deterministic for a given combination of input parameter values, and GSA requires running CICE many times (~1000 per input parameter, [*Oakley and O'Hagan*, 2004]) to fully explore the parameter space, which is impractical from a computational standpoint. An alternative is to build a fast surrogate (or emulator) of CICE to produce estimates for any combination of input parameters. We use a Bayesian approach to represent scalar output variables of interest (sea ice extent, area, and volume) in terms of

unknown functions following Gaussian distributions at every input point (i.e., every combination of 39 input parameters). That is, the prediction of the emulator has a Gaussian distribution ($\mathcal{N}$) which is fully characterized by its mean and covariance. An underlying assumption of Gaussian process emulators is that the model response is a smooth function of the inputs, so that nearby input points have similar response function values. A Gaussian process of a model output variable $Y = f(\mathbf{X})$ is described as

$$f(\cdot) \sim \mathrm{GP}(\mu(\cdot), K(\cdot,\cdot)) \quad (6)$$

where $\mu(\cdot)$ is the mean function, and $K(\cdot,\cdot)$ the covariance function specifying the covariance between pairs of random variables. Prior assumptions about the mean and covariance functions must be made and then updated using evidence from a training dataset. First, the training and test points are given by $\mathbf{X_t}$ and $\mathbf{X}^*$, respectively, and the mean and covariance priors by μ^* and $K(\mathbf{X}^*, \mathbf{X})$, respectively. Without loss of generality the prior mean (μ^*) is typically assumed to be zero. Secondly, in order to obtain predictions of $f(\cdot)$, the prior is updated using evidence coming from the training dataset. The distribution of $f(\mathbf{X}^*)$ conditioned on the training data $f(\mathbf{X_t})$ is then (see *Rasmussen and Williams* [2006] for details on Gaussian processes)

$$f(\mathbf{x}^*)|f(\mathbf{x_t}) \sim \mathcal{N}(\mu^* + K(\mathbf{x}^*, \mathbf{x_t}) \cdot K(\mathbf{x_t}, \mathbf{x_t})^{-1} \cdot (f(x_t) - \mu_t), K(\mathbf{x}^*, \mathbf{x}^*) - K(\mathbf{x}^*, \mathbf{x_t}) \cdot K(\mathbf{x_t}, \mathbf{x_t})^{-1} \cdot K(\mathbf{x_t}, \mathbf{x}^*)) \quad (7)$$

The form of the covariance between predictions in terms of the input parameters requires covariance functions, which have a number of hyperparameters controlling the smoothness (length scales) and the overall variance (signal variance) of the functions. The covariance structure is separable in parameter space, implying that length scales are different for each parameter. We use the GPy library implemented in Python [*The GPy authors*, 2012-2015] to fit emulators for each scalar output variable of interest (i.e. sea ice extent, area, and volume). Emulators were fit using different covariance functions including exponential, squared exponential, Matern 3/2, Matern 5/2, and combinations of sums and products among them [*Rasmussen and Williams*, 2006]. Multiplication and summation of covariance functions can, in principle, improve emulator performance when the process underlying a dataset has more than

one length scale associated with it. Each time an emulator is fitted, the hyperparameters are optimized by maximizing the likelihood of the data given the fitted emulator (marginal likelihood). In addition, logarithmic and square root transformations were also applied during the fitting process to obtain better emulation performance [*O'Hagan*, 2006].

We have followed a cross-validation approach to choose among emulators with different covariance functions and different data transformations. In this context, the entire set of actual model runs is used for training and validation purposes. In this study the entire dataset is divided into 10 subsets. Each subset is used as validation dataset once, while the remaining 9 subsets are used as training datasets. The metric used to assess the emulator performance is the square error between the emulated and simulated mean quantities. For each covariance function and data transformation, a root-mean-square error (RMSE) is computed over all the validation points in the 10 emulators fitted during cross validation, and the emulator combining the covariance function and data transformation with the smallest RMSE is chosen for a particular output variable of interest (section 4.2). The entire dataset (as opposed as subsets) is then used to train an emulator that uses the optimal combination of covariance function and data transformations as described above.

4. Results

In this section we present results of sea ice extent, area and volume from an ensemble of 302 members to illustrate the propagation of parametric uncertainty onto model predictions. We also present results from the cross-validation of Gaussian emulators (built using an ensemble of 397 members) for the output variables of interest. At the end of the section we present the sensitivity indices for the CICE model, including a ranking of the most important parameters contributing to the CICE model uncertainty in the current model configuration.

4.1. Parameter uncertainty propagation onto CICE predictions

We use the results of 302 model runs whose configurations used the parameters prior distributions (~24% of runs crashed due to instabilities caused by some parameter combinations) to examine how uncertainty in the input parameters (Table 2) propagates onto model predictions

(i.e. sea ice extent, area, and volume). Figure 1 shows histograms of sea ice extent, area, and volume produced by the ensemble of 302 simulations in the northern hemisphere in September and March. In summer the spread is considerably larger and the variables are more uniformly distributed than in winter. The winter estimates have less spread and are characterized by more unimodal distributions than in summer, indicating that parameter uncertainty is less important in winter. In the southern hemisphere (Figure 2), the cross-ensemble variability is relatively small and similar in March and September for sea ice extent and area. This variability is, however, much larger in September than in March for sea ice volume, indicating that ice thickness can be very sensitive to parameter choices during winter in the southern hemisphere.

The propagation of parametric uncertainty onto the monthly mean extent, area and volume is shown in Figure 3. After the maximum ice extent is reached in March in the northern hemisphere, parametric uncertainty in the sea ice extent and area estimates begins to grow, becoming considerably large by the end of spring through the end of summer (Figure 3a, c). With ice starting to grow in fall, uncertainty in extent and area is reduced, becoming almost negligible throughout the winter. The low uncertainty in the winter sea ice area estimates is explained by the prescribed atmospheric forcing driving high concentrations over most of the Arctic Ocean. The winter sea ice extent estimates, however, are more affected by the melting rates at the ice-ocean interface, which are determined largely by the prescribed climatological ocean conditions in standalone simulations of sea ice [*Bitz et al*., 2005]. Unlike extent and area, parametric uncertainty in sea ice volume is large all year round in the northern hemisphere (Figure 3e). It is expected that in a fully couple model parametric uncertainties of sea ice extent and area also become important all year round (including winter). In a fully coupled context the choice of sea ice parameters should feedback on the atmosphere, which would result in uncertainty estimates of winter sea ice extent and area larger than those in Figure 3a and c.

In the southern hemisphere the uncertainty in the model estimates of sea ice extent and area is the lowest during fall, when sea ice begins to grow, and during winter. Once the sea ice begins to retreat in spring, the uncertainty rapidly increases. With the sea ice pack coverage dramatically shrinking in summer, most of the ensemble runs become insensitive to parametric uncertainty. The uncertainty in the model estimates of sea ice volume in the southern hemisphere has a more

distinctive seasonality than in the northern hemisphere. As the sea ice pack starts to shrink in the Southern ocean during spring, the uncertainty in sea ice volume estimates begins to decay. With the expansion of the sea ice pack in winter, the volume estimates have large uncertainty and the choice of parameter values becomes very important.

The CICE default configuration typically produces estimates within one standard deviation above the ensemble mean estimates. Observational estimates shown in Figure 3 are generally within model ensemble estimates, except for a slight overestimation of the sea ice area in the ensemble during winter in both the northern and the southern hemisphere. Further reduction in the uncertainty of input parameters are not expected to improve considerably the agreement in the annual cycle between model estimates and observations because their difference is several times bigger than the range of model predictions. However, uncertainties associated with the ocean and atmospheric forcing can be more relevant than parametric uncertainty to produce better model agreement with observations, particularly when parametric uncertainty is considerably low.

4.2. Emulator validation

From the preliminary screening (section 3.4) we expect around 10 active parameters per output variable. Since the required number of model runs required to build an emulator is recommended to be at least 10-15 per active variable [*O'Hagan, 2006; Oakley and O'Hagan, 2001*], we have conducted a total of 550 model runs as described by the experimental design (section 3.3). The mapping of the Sobol' design onto parameter space was made as described in section 3.2 using a mixture of the parameters prior distributions and uniform distributions.

Using the monthly sea ice extent, area and volume from an ensemble of 397 successful model runs (~28% runs failed out of 550) we fit Gaussian emulators and select the one which minimizes the RMSE between training points and test points. The emulator with the best performance is then used to conduct the sensitivity analysis. Figure 4 shows the RMSE values obtained for the emulators for which the best fit is obtained as described in section 3.5. The accuracy of the emulator in terms of RMSE varies throughout the year, which is closely related to the variable uncertainty found in the ensemble runs as indicated in Figure 3.

In the northern hemisphere the largest errors in the emulated extent and area occur in August (Figure 4), when the parametric uncertainty in the ensemble is also the largest. The RMSE in summer area and extent is about 10-20% of the ensemble mean quantities. As expected from Figure 3, the emulator performance in winter, spring, and fall is very good. The emulator errors for sea ice volume in the northern hemisphere are relatively constant throughout the year in accordance with the uncertainty in the ensemble, and the RMSE as a percentage of the ensemble mean thickness decreases from about 17% in September to about 6% in April.

In the southern hemisphere the RMSE is generally smaller than in the northern hemisphere. The emulated sea ice extent, area, and volume have relatively better accuracy from February-October than from November-January. The RMSE as a percentage of the ensemble mean quantities has a maximum value of about 20% (February), and values less than 2% from November-January indicating good emulator accuracy for most of the year in the southern hemisphere. The relatively smaller RMSE in southern hemisphere than in the northern hemisphere suggests that the emulator generally produces better estimates in seasonal ice than in multiyear ice regimes.

We show comparisons of emulated and simulated September sea ice volume conducted during cross-validation over different validation sets for the northern (Figure 5) and the southern hemispheres (Figure 6). In the northern hemisphere most of the model simulations fall within one standard deviation of the mean emulation prediction. There are a few validation points where the disagreement between emulated and simulated sea ice volume is relatively important, particularly in the validation set 9, possibly due to parameter subspaces in which sampling of training points is not optimal due to rapid variations of the model predictions. In the southern hemisphere the emulator predictions match very well the model estimates in all the validation sets.

The emulator has shown skill in predicting model results within reasonable error bands. The least agreement is found at the end of summer in the northern hemisphere. Given the complexity of the processes controlling summer Arctic sea ice variability, the emulator performance is still satisfactory for the purpose of conducting the sensitivity analysis, whose results are presented in the next section.

4.3. CICE sensitivity indices

In GSA, total sensitivity indices give the overall importance that input parameters have on model sensitivity; parameters with total sensitivity indices close to zero are unimportant in terms of sensitivity and can be fixed at some default value without significantly affecting the model predictions. Parameters with low first order sensitivity indices could still have important interactions with other parameters and therefore should not be set at fixed values.

The first order and total sensitivity indices are estimated as described in the Appendix using N=10^5 to guarantee the convergence of sensitivity indices with p=39 parameters. This implies that the required number of emulator estimates per output variable of interest is 2×10^5 emulator runs (**A** and **B** matrices) plus 10^5 ($\mathbf{A_B^i}$ matrix) emulator runs per input parameter. Therefore a total of 4.1×10^6 emulation predictions are required to obtain the sensitivity indices of a single output variable.

The seasonal evolution of the total sensitivity indices in the northern hemisphere is shown in Figure 7 for the sea ice extent, area and volume for a selection of the parameters contributing to the model uncertainty (see also Tables 1 and 2 for parameter names and ranges of variation). The model predictions are highly sensitive to the snow conductivity (*ksno*) all year round with $S_T >$ 0.47 for ice volume and $S_T > 0.2$ for extent and area. Parameters controlling the melt pond drainage (*lambda_pond*) and the maximum snow grain size (*rsnw_mlt*) are important during summer and early fall with $S_T > 0.15$. The snow grain size also affects model predictions through another parameter, *R_snw*, with $S_T > 0.05$ from spring to fall. The ocean-ice drag is very important for sea ice extent and area through the *dragio* parameter during winter and spring when sea ice is advected less easily under the effect of seasonal winds, and considerably less important during summer and early fall when the ice drifts more freely. It is possible that the effect of *dragio* on sea ice volume is less important than on extent and area because volume is more easily affected by thermodynamics than by dynamical processes such as redistribution and opening of leads, particularly when the ice pack is tight. The sea ice area at the end of fall and beginning of winter is also affected by the assumed values of the solid fraction at the ice-ocean interface (*phi_i_mushy*) in the mushy thermodynamics formulation. Generally, as the total index (S_T) of an individual parameter is reduced, the total indices of one or more other parameters

increase, indicating competing effects among relevant physical processes affecting sea ice throughout the year. For instance, the sensitivity of the sea ice volume to *ksno* drops considerably at the end of spring and remains relatively low through summer in favor of the sensitivity to *lambda_pond*, *rsnw_mlt*, and *R_snw*.

At the onset of the melt season, the sea ice and snow layers begin to melt and absorb solar radiation, making the latter parameters become active. The extent to which melt ponds are drained has an effect on the albedo of sea ice, which is a combination of the albedos of bare and ponded sea ice. The larger the fractional area of ponded ice, the more heat is absorbed by the ice pack, which has a direct effect on the melting of sea ice. Melt ponds can be formed by snow melt runoff, which depends also on the heat absorption of the snow pack. The snow albedo in the delta-Eddington formulation for absorption of shortwave radiation depends, among other things, on the snow grain size. This sensitivity analysis shows that *rsnw_mlt* and *R_snw* are the delta-Eddington parameters with the highest sensitivity in sea ice model predictions.

Sea ice in the southern hemisphere (Figure 8) is also highly sensitive to snow conductivity (*ksno*), though to a lesser extent than in the northern hemisphere. In the southern hemisphere dynamical processes acquire a larger relevance through processes involving *dragio* in the ocean-ice drag formulation. In contrast to the northern hemisphere, the ocean-ice drag is important not only in winter and spring but also in summer. The *R_snw* and *rsnw_mlt* parameters affecting the snow grain size are still somewhat important in the southern hemisphere although to a lesser degree than in the northern hemisphere.

Even though sensitivities of sea ice extent and area are generally correlated with each other, there are instances when the effect of one parameter on sea ice area differs considerably from the effect on sea ice extent. For instance, Figure 7b shows an increase of sensitivity of sea ice area to *dragio* while sea ice extent remains insensitive to that parameter in May. This spike in the sensitivity of sea ice area to *dragio* (Figure 7b) is accompanied by a sudden drop in the sensitivity to *ksno* (Figure 7a). From March to May when ice concentrations are relatively high, the changes in concentration will affect more the area than the extent. The sensitivities on sea ice area and extent could also differ due to changes in concentrations taking place away from the marginal ice zone, to changes in sensitivities associated with other parameters, or due to errors in

the sensitivity estimates arising from emulator error.

In Figure 9 and Figure 10 we present the first order and the total sensitivity indices of the CICE input parameters for sea ice at the end of each of the four seasons. We focus the analysis on sea ice volume because it is an integrated measure of the thickness, which to a large extent determines the survivability of sea ice and therefore can be a more relevant variable to predict sea ice evolution [*Holland et al.*, 2008; *Chevallier and Salas-Mélia*, 2012]. We have ranked the indices in order of decreasing importance for the September ice volume (month of minimum Arctic sea ice coverage) in the northern hemisphere. In the figures we have used 0.02 (shaded region) as a threshold to indicate that an effect is unimportant. The difference between the total sensitivity and the first order indices for an individual parameter along the horizontal axis indicates the relative importance of its interactions with other parameters. If the total sensitivity index is close to the first order index for a given parameter, there are no important interactions with other terms, and the effect of that parameter on the model predictions is purely additive (i.e., it does not depend on what values other parameters take). If the total sensitivity index is not close to the first order index however, then the effect of an individual parameter is non-additive and will depend on the values that interacting parameters take.

The September sea ice volume in the northern hemisphere (Figure 9a) is sensitive to about 10 parameters. The top 5 most important parameters are *ksno*, *lambda_pond*, *rsnw_mlt*, *R_snw*, and *dragio*, all of which display interactions with other parameters. The other parameters with non-negligible sensitivity indices include density of sea ice and snow (*rhoi* and *rhos*, respectively), the thickness of the ice scattering layer (*hi_ssl*), the aspect ratio of the yield curve (*ecc*), and the strength of the slow drainage mode in the mushy thermodynamics (*dSdt_slow_mode*). In fall and winter (Figure 9b and c) the sensitivity of the model becomes larger for *ksno* but reduces for the other parameters. The high sensitivity to *ksno* highlights the importance of the insulating effect of the snow cover (which will be discussed in section 5) along with the effects that other parameters have on sea ice volume. During the ice growth season the interaction between model parameters also decreases and sensitivity is due mainly to first order effects. In spring, the *R_snw* indices begin to spin up (Figure 9d), but other parameters that are also active during the melt season only become important in summer (Figure 9a).

In the southern hemisphere the sea ice volume is also highly sensitive to *ksno*. In comparison with the northern hemisphere, however, the importance of *lambda_pond* and *rsnw_mlt* is considerably reduced. During summer (Figure 10c), *dragio*, *rhoi,* and *ecc* are considerably more important in the southern hemisphere than they are in the northern hemisphere. The energy dissipation parameter due to shear (*Cs*) is important only in the southern hemisphere. Without considering the effect of *ksno*, this would indicate that in comparison with the Arctic, the sensitivity of sea ice in the southern ocean appears to be more affected by dynamical processes than by thermodynamic processes. The interactions between model parameters are also less important in the southern hemisphere, except for interactions associated with *dragio* which appear to be produced almost entirely by interaction with *ksno* (Figure 10c). In fall, winter, and spring (Figure 10d, a, and b), sea ice volume is sensitive to fewer parameters and their interactions are small. Besides sensitivity to *ksno*, important parameters also include *dSdt_slow_mode* in fall and winter, and *rhoi* and *astar* in winter and spring.

We also estimated the sensitivity indices in different ice regimes, one during the 1980s when relatively large amounts of sea ice were present in the Arctic, and another during the 2000s when historical minima of sea ice occurred in the Arctic (not shown). There was no evidence that the sensitivity to input parameters changed considerably between these two different climatic conditions, with only marginal differences in the fractions of the total variance produced by individual parameters, and shifts in the parameter rankings in only a few cases.

Our parameter rankings are broadly consistent with previous sensitivity studies. In particular, ice and snow conductivities, and albedos have been found to be the most important parameters in *Kim et al.* [2006], *Peterson et al.* [2010], and *Uotila et al.* [2012]. While our study does not explicitly examine albedos because these are computed internally within CICE, it identifies important parameters controlling snow grain size (*rsnw_mlt* and *R_snw*) which ultimately affect the albedo estimates in the delta-Eddington radiation scheme. Our study also identifies important parameters previously found unimportant or not examined such as *lambda_pond*, *rsnw_mlt*, *astar*, and *hi_ssl.* These parameters suggest the important role of specific melt pond (drainage), snow (grain size) and ice (scattering, and ridging) parameters contributing to model uncertainty. Arctic ice volume is relatively less sensitive to ice density (*rhoi*) in this study than in previous

studies [*Kim, et al.*, 2006] because its importance is superseded by the snow and melt pond parameters mentioned above. A comparison with a previous sensitivity study in a coupled climate model [Rae et al., 2014] shows that conductivity is the most important parameter in both standalone and fully coupled configurations. While *R_snw* is important in both configurations, our study suggests that other parameters (*rsnw_mlt*, *lambda_pond*) are even more important, which could be partially explained by the more comprehensive list of input parameters included in our study.

5. Discussion of main effects and interactions

The sensitivity indices estimated in section 4.3 measure the sensitivity of sea ice to the individual parameters and the interactions among them. They do not, however, provide much insight onto the functional relationships between the sea ice variables and the effects of individual parameters, or the physical processes underlying such relationships. We focus on the September Arctic sea ice volume to discuss its relationships with the most important parameters in terms of parametric uncertainty. We seek a generalized additive model [*Hastie and Tibshirani, 1986*; *Wood, 2000*] with the structure given by equation (1), but with inputs (X_i) restricted to the top ten ($p=10$) most influential parameters in Figure 9a. A generalized additive model assumes that the model prediction is the sum of unknown functions of the parameters, which are estimated non-parametrically rather than based on assumed forms such as polynomials. For simplicity we consider only the first and second order terms in equation (1), and neglect higher order interactions so that $f(\mathbf{X}) \approx f_0 + \sum_{i=1}^{10} f_i(X_i) + \sum_{1\le i<j}^{10} f_{ij}(X_i, X_j)$. The main effects are the univariate functions in the first summation, and the second order effects (or interactions) are the bivariate functions in the second summation of the expression for $f(\mathbf{X})$. In the expression for $f(\mathbf{X})$ the expected value over all the input parameters is f_0 and therefore the main and the second order effects represent deviations from that expected value.

In order to fit an additive model for September Arctic sea ice volume we evaluate the emulator 10^4 times, varying the top 10 important parameters and setting the unimportant parameters fixed at default values. The main effects (one per parameter) are obtained by applying non-parametric smoothing functions to one-dimensional scatterplots of sea ice volume

versus each of the 10 parameters included in the model [*Wood, 2006; Wood, 2015*]. Similarly, the interactions between pairs of parameters are obtained by smoothing two-dimensional scatterplots of the sea ice volume as a function of pairs of parameters, which amount to 44 second order interactions.

5.1. Main effects

Figure 11 shows the main effects on the most important model parameters controlling the simulated September Arctic sea ice volume. The main effects give the sensitivity of the model to variations in one individual parameter when the model predictions are averaged over all other parameters. Thus the figure shows first order changes in sea ice volume as a function of individual parameters along their (normalized) plausible range of values. Note that the relationships between ice volume and *lambda_pond*, *rsnw_mlt*, *dragio*, *ecc*, and *ksno* have clear non-linear behavior.

The variation of sea ice volume is larger as a function of *ksno* (Figure 11a) in comparison with the variation produced by other parameters (Figure 11b-i). An increase in snow conductivity reduces the insulation effect of the snow, resulting in larger heat losses and ice growth during winter. As *lambda_pond* increases melt pond drainage occurs faster, reducing the fractional area of ponded sea ice. Ponded sea ice has lower albedo than bare ice and therefore promotes sea ice melt. Large values of *lambda_pond* associated with bare ice lead to large sea ice volume while lower values associated with ponded ice result in reduced volume (Figure 11b). The main effect for *lambda_pond* tends to flatten near the extremes, which can be indicative of thresholds beyond which sea ice volume is insensitive. These thresholds could occur because virtually all sea ice is ponded (low *lambda_pond*), or because melt ponds have been totally flushed out of the sea ice surface (high *lambda_pond*).

The sensitivity of the model to parameters involved in the radiation scheme is encompassed mainly in *rsnw_mlt*, *R_snw*, and *hi_ssl* (Figure 11c, d, and g). The first two parameters affect directly the estimation of the snow grain size, and the third parameter is the thickness of the ice scattering layer. For relatively small snow grain sizes, the maximum melting snow grain size (*rsnw_mlt*) has a big influence on sea ice volume (Figure 11c). Constraining the snow grain size

also constrains the lower bound of snow albedo. When *rsnw_mlt* is small, increasing its value results in a dramatic decline of ice volume due to considerable reduction in snow albedo and enhanced absorption of solar radiation. The sensitivity of sea ice volume is much less important at high values of *rsnw_mlt* possibly due to high melt rates which would deplete the snow pack beyond certain grain sizes. The *R_snw* parameter gives the standard deviation of the snow grain size. Increasing *R_snw* has the effect of reducing the grain size in the delta-Eddington radiation scheme, increasing albedo and reducing absorption of solar radiation, which results in lower sea ice melt rates and increased September ice volume (Figure 11d). Increasing the thickness of the ice scattering layer (*hi_ssl*) considerably increases the ice volume (Figure 11g) by increasing the albedo and reducing the heat absorption of bare ice [*Light et al., 2015*], thus reducing the overall melting in the sea ice pack.

September Arctic sea ice volume can also be sensitive to dynamic processes in the standalone model through the *dragio* and *ecc* parameters (Figure 11e and i). Sea ice volume is slightly sensitive at low values of the neutral ocean-ice drag, but largely sensitive at high values. Moreover, the main effect of *dragio* suggests that sea ice volume can either increase or decrease with *dragio*. At low *dragio* values, increasing of the neutral drag results in a slight reduction of Arctic sea ice volume, possibly because of the dominant role of the wind stress that would keep moving sea ice out of the Arctic into the North Atlantic. On the other hand, at relatively high values of *dragio*, increasing the neutral drag increases the ice volume. We speculate that enhancing the ocean-ice drag could slow down the transpolar drift, reducing the export of sea ice through Fram Strait and accumulating sea ice in the central Arctic. This would be consistent with previous studies suggesting that Arctic ice thickness is affected by Fram Strait sea ice export [*Langehaug et al.,* 2013]. Sea ice volume is also sensitive to the aspect ratio of the elliptical yield curve (*ecc*). In Figure 11i, sea ice volume decreases as *ecc* increases. Increasing *ecc* diminishes the capacity of the ice pack to withstand shear stress and the reduction in ice volume results from increased ice velocities and larger exports by the transpolar drift.

The simulated sea ice volume is also sensitive to the choice of sea ice and snow densities (*rhoi* and *rhos*, respectively). Their main effects in Figure 11f and h show that the September Arctic sea volume monotonically increases as both *rhoi* and *rhos* increase. These parameters are

used in both dynamic and thermodynamic modules and affect CICE model estimates in different ways. Figure 11h would suggest that the predominant effect of the snow density on September Arctic sea ice volume is reduced insulation in winter (and increased ice growth) due to the thinning of snow layers as snow density increases. Finally, Figure 11f shows that increasing *rhoi* results in higher Arctic ice volume, consistent with increasing compressive and shear strength of the ice pack [*Rothrock, 1975*; *Liscomb et al., 2007*]. The transpolar drift slows down under relatively more resistant sea ice pack, and subsequent accumulation of sea ice occurs in the central Arctic.

5.2. Interactions

We present only the three most important interactions (out of 44) based on *F*-statistics obtained upon fitting the additive model derived from equation (1). These correspond to the interactions of *ksno* with *lambda_pond*, *rsnw_mlt*, and *R_snw*. Figure 12 shows these interactions as bivariate functions accounting for the fact that the effect of one parameter on the model predictions might depend on the value of another parameter. Interaction in this context means that the combined effect of interacting parameters is different from the sum of their main effects. The interactions can be thought of as "effect modification" terms to be applied to the main effects, depending on the value of the interacting parameter.

Figure 12a shows the second order interaction between *ksno* and *lambda_pond*. At low values of *lambda_pond*, the sea ice volume decreases as *ksno* increases, which has the effect of flattening the main effect of *ksno* in Figure 11a. At high values of *lambda_pond* the second order interaction along *ksno* has the opposite effect, and has the effect of steepening the main effect of *ksno*. The effect of *ksno* on the sea ice volume, up to the second order, is then more pronounced in the absence of melt ponds (high *lambda_pond*) than in a scenario with large amounts of melt ponds (low *lambda_pond*). Clearly, as melt ponds become more ubiquitous over the ice pack, sea ice melting becomes more important due to increased absorption of solar radiation. The fraction of sea ice covered by snow decreases and so does the overall effect of heat conduction through the snow pack. The *ksno-lambda_pond* interaction (Figure 12a) also modifies the main effect of *lambda_pond* on sea ice volume in Figure 11b, indicating that at relatively low values of *ksno,* the ice volume decreases as *lambda_pond* increases, which would flatten the main effect of

lambda_pond in Figure 11b. At high values of *ksno*, the interaction is an increasing function of *lamba_pond*, which would steepen the main effect of *lambda_pond* in Figure 11b. The steepening of the main effect indicates that sensitivity to *lambda_pond* increases, and vice versa. When the conductivity of the snow pack (*ksno*) is high there is considerable sea ice coverage, which could sustain large coverage by melt ponds, which in turn would have a large effect on melt rates and volume changes. If *ksno* is relatively low, the ice pack shrinks considerably and it is easy for melt ponds to become ubiquitous on the ice pack, where any addition of ponds has no further effect on the ice volume.

Another important interaction occurs between *ksno* and *rsnw_mlt*. Low values of *ksno* in Figure 12b tend to flatten the main effect of *rsnw_mlt* in Figure 11c, and high values of *ksno*, to steepen the main effect of *rsnw_mlt*. The sensitivity of ice volume to ponded ice is large when there is large sea ice coverage, and ponds appear more quickly as the maximum melting snow grain size (*rsnw_mlt*) increases, because the snow albedo is less constrained towards high values. If on the other hand, the overall amount of sea ice is relatively low (low *ksno*), the surface of the ice pack quickly fills with ponds so that increasing *rsnw_mlt* no longer changes considerably the ponded ice albedo nor the total sea ice volume. When the maximum melting snow grain size (*rsnw_mlt*) is small, the *ksno-rsnw_mlt* interaction (Figure 12b) tends to steepen the main effect of *ksno* in Figure 11a. Low *rsnw_mlt* values limit the amount of solar radiation absorbed by the snow pack resulting in relatively less ponded ice, which enhances the sensitivity to heat conduction through snow layers. The opposite effect takes place as the main effect of *ksno* is flattened by the interaction of *ksno-rsnw_mlt* at relatively large snow maximum melting grain sizes.

Similar to the *ksno-rsnw_mlt,* the interaction between *ksno-R_snw* also indicates an interplay between snow grain size and snow conductivity (Figure 12c). Both *rsnw_mlt* and *R_snw* affect directly the snow grain size and therefore the snow albedo. Note, however, that by increasing *R_snw* the snow grain size decreases as parameterized in the delta-Eddington formulation for shortwave radiation. Thus at high values of *ksno*, the interaction between *ksno* and *R_snw* increases the sensitivity of the main effect of *R_snw*, and vice versa (Figure 12c and Figure 11d).

Again, as with the interaction between *ksno-rsnw_mlt*, smaller grain sizes (or large *R_snw*) increase the sensitivity of the model to snow conductivity (*ksno*).

6. Summary and conclusions

This study presents a new approach for quantifying uncertainty in complex sea ice models that shows promise for applications in the context of fully coupled climate models. We have conducted a variance-based global sensitivity analysis of the Los Alamos sea ice model (CICE) to quantify the sensitivity of simulated sea ice conditions to input parameters in the model. This approach explores the sensitivity of the model over the full range of uncertainty of 39 input parameters. In the method used here there are no implicit assumptions on linearity or additivity of the model. We used a Gaussian process emulator to approximate model estimates of sea ice extent, area and volume. The emulation approach allowed numerous estimates of model predictions which are required to estimate key sensitivity measurements, the first order and the total sensitivity indices. By computing these effects we were able to identify the model parameters whose uncertainty contribute considerably to the model uncertainty, and elucidate the importance of interactions among input parameters. The estimation of the sensitivity indices are quantities to be used for research prioritization and establishing what parameters could be fixed to default values without significantly affecting the sensitivity of the model predictions. It is recommended that further research efforts are made to acquire more knowledge about the accurate value of parameters with high first order sensitivity indices, including calibration activities.

The main physical result of this study is the identification of sea ice processes and parameters to which the CICE model is most sensitive. These include snow conductivity (*ksno*), drainage of melt ponds (*lambda_pond*), snow grain size (involving *rsnw_mlt*, *R_snw*), the thickness of the ice scattering layer (*hi_ssl*), snow and ice densities (*rhos* and *rhoi,* respectively), the ice-ocean drag (involving *dragio*), and the aspect ratio of the yield curve (*ecc*). Their effects on sea ice predictions were shown to vary throughout the year indicating seasonality of the main processes affecting sea ice extent, area, and volume. It was also shown that in comparison with the southern hemisphere, there are more active parameters driving the sensitivity of the model in the

northern hemisphere. Interactions among parameters are also more important in the northern hemisphere than in the southern hemisphere.

We also interpreted the main processes driving the sensitivity of the model by interpreting the functional relationships between September Arctic sea ice volume and the most important parameters according to the sensitivity indices. Several of the main effects are non-linear relationships, in which the sea ice sensitivity to an individual parameter depends on the value of the parameter itself (e.g. drainage of melt ponds, ice-ocean drag, or the maximum snow grain size). The analysis also identified the most important interactions among input parameters, which have the effect of modifying the main effects of individual parameters, depending on what value the other interacting parameter takes. We found that important interactions between the snow conductivity (*ksno*) and the drainage of melt ponds (*lambda_pond*), and between the snow conductivity and the snow grain size (*rsnw_mlt*, *R_snw*). Previous studies had already identified the large sensitivity of sea ice variables to snow conductivity, but its interactions with the snow and melt pond parameters had not yet been identified. These interactions indicate an interplay between the conduction of heat through the snow pack and its effect on the ice volume, and the relative amounts of ponded sea ice which affect the albedo and melting rates of sea ice. It follows that interactions among parameters can be important, and that sensitivity studies in climate applications must consider these types of effects because additivity of main effects may not be justified.

This study also found that model estimates of sea ice volume generally display more parametric uncertainty than estimates of sea ice extent and area. Model estimates have more uncertainty associated with parameters during summer than in winter. In particular, fall and winter estimates of sea ice area and extent have very little parametric uncertainty in both the northern and the southern hemisphere. Model predictions of sea ice extent and area during these seasons are expected to be affected more by atmospheric and oceanic forcing, than by parametric uncertainty. Summer predictions of sea ice conditions are largely affected by parametric uncertainty, and efforts should be made to reduce this source of uncertainty to achieve more robust prediction systems.

We acknowledge the fact that the sensitivity of a coupled atmosphere-ocean-ice system might differ from the results obtained in this study because of feedbacks among these components, which can affect wind patterns and heat transport through changes in clouds and atmospheric boundary layer [Taylor et al., 2013; Rae et al., 2014]. We will examine this issue for a reduced number of model parameters in a future study. We also point out that bringing more complexity to climate models requires a good characterization of new uncertainties incorporated in the modeling framework, as well as a complete assessment of the model sensitivity to new parameters included by new developments. This study has addressed this issue for the latest version of the CICE sea ice model, which includes numerous new physical parameterizations.

Acknowledgements

We want to thank Andrew Roberts and two anonymous reviewers for useful discussions and suggestions on the manuscript. This study has been supported by the Regional and Global Climate Modeling (RGCM) Program of the Office of Biological and Environmental Research (BER) within the U.S. Department of Energy's Office of Science. We thank the Earth System Modeling (ESM) Program, also within BER, for use of the column-package version of CICE5, as developed for the Accelerated Climate Model for Energy (ACME). The sea ice observational data used in this study was obtained freely from the National Snow and Ice data Center and from papers properly cited and referred to in the reference list. The source code for CICE 5.1 and other data are available from the authors upon request at jorge.urrego.blanco@lanl.gov.

Appendix. Estimation of sensitivity indices

The direct estimation of the sensitivity indices (section 2) is made using the mean predictions produced by the emulator. Evaluation of equations (4) and (5) involves evaluation of integrals which is typically done using Monte Carlo, or quasi-Monte Carlo numerical integration. In practice, several estimators have been derived in previous studies [*Sobol'*, 2001; *Saltelli et al*., 2010; *Owen*, 2013]. We have followed [*Saltelli et al*., 2010] and used the following estimators for the sensitivity indices in terms of emulations estimates at sampling matrices **A**, **B**, and $\mathbf{A_B^i}$, each containing *N* emulator runs.

$$V_{\mathbf{X}_i}\left(E_{\mathbf{X}_{\sim i}}(Y|X_i)\right) \approx \frac{1}{N}\sum_{j=1}^{N} f(\mathbf{B})_j\left(f\left(\mathbf{A_B^i}\right)_j - f(\mathbf{A})_j\right) \quad \text{(A1)}$$

and

$$E_{\mathbf{X}_{\sim i}}\left(V_{\mathbf{X}_i}(Y|X_{\sim i})\right) \approx \frac{1}{2N}\sum_{j=1}^{N}\left(f(\mathbf{A})_j - f(\mathbf{A_B^i})_j\right)^2 \quad \text{(A2)}$$

These matrices have dimensions $N \times p$ corresponding to *N* emulator predictions and *p* parameters. The matrices **A** and **B** are Sobol' sequences with *N* sample points. To obtain $\mathbf{A_B^i}$**,** for $\{i = 1, \ldots, p\}$, the i^{th} column of matrix **A** is replaced by the column i^{th} of **B**. To obtain the sensitivity indices, equations (A1) and (A2) need to be normalized by the total variance *V(Y)* which is estimated as the variance over all the emulation predictions of **A** and **B**.

References

Alexandrov, V., S. Sandven, J. Wahlin, and O. M. Johannessen (2010), The relation between sea ice thickness and freeboard in the Arctic,*e,* 4, 373-380.

Andreas, E. A., P. O. G. Persson, R. E. Jordan, T. W. Horst, P. S. Guest, A. A. Grachev, and C. W. Fairall (2010), Parameterizing turbulent exchange over sea ice in winter, *J. Hydrometeorol.,* 11, 87-104.

Andreas, E. L., T. W. Horst, A. A. Grachev, P. O. G. Persson, A. W. Fairall, P. S. Guest, and R. E. Jordan (2010), Parameterizing turbulent exchange over summer sea ice and the marginal ice zone, *Q. J. R. Meteorolo. Soc.,* 136, 927-943.

Aoki, T., T. Aoki, M. Fukabori, A. Hachikubo, Y. Tachibana, and F. Nishio (2000), Effects of snow physical parameters on spectral albedo and bidirectional reflectance of snow surface, *J. Geophys. Res.,* 105(D8), 10219-10236.

Bengio, Y., O. Delalleau, and N. Le Roux (2006), The curse of highly variable functions for local kernel machines. In : *Advances in neural information processing systems*. MIT Press.

Bentsen, M., I. Bethke, J.B. Debernard, T. Iversen, A. Kirkevåg, Ø. Seland, H. Drange, C. Roelandt, I.A. Seierstad, C. Hoose, and J.E. Kristjánsson (2013), The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate, *Geosci. Model Dev.*, 6, 687-720, doi:10.5194/gmd-6-687-2013.

Bitz, C.M., M.M. Holland, E.C. Hunke, and R.E. Moritz (2005), Maintenance of the sea ice edge, *J. Climate*, 18, 2903-2921.

Blockley, E.W., M.J. Martin, A.J. McLaren, A.G. Ryan, J. Waters, D.J. Lea, I. Mirouze, K.A. Peterson, A. Sellar, and D. Storkey (2014), Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts, *Geosci. Model Dev.,* 7, 2613-2638, doi:10.5194/gmd-7-2613-2014.

Brandt, R. E., S. G. Warren, A. P. Worby, and T. C. Grenfell (2005), Surface albedo of the Antarctic Sea Ice Zone, *J. Climate,* 18, 3606-3622.

Briegleb, B. P., and B. Light (2007), *A Delta-Eddington multiple scattering parameterization for solar radiation in the sea ice component of the Community Climate System Model.* Technical Note, National Center for Atmospheric Research.

Chevallier, M., and D. Salas-Mélia (2012), The role of sea ice thickness distribution in the Arctic sea ice potential predictability: A diagnostic approach with a coupled GCM, *J. Climate*, 25,

3025-3038, doi: 10.1175/JCLI-D-11-00209.1.

Collins, W. D., C. M. Bitz, M. L. Blackmon, G. B. Bonan, C. S. Bretherton, J. A. Carton, P. Chang, S. C. Doney, J. J. Hack, T. B. Henderson, J. T. Kiehl, W. G. Large, D. S. McKenna, B. D. Santer, and R. D. Smith (2006), The community climate system model version 3 (CCSM3), *J. Climate,* 19, 2122-2143.

Derksen, C., J. Lemmetyinen, P. Toose, A. Silis, J. Pullianinen, and M. Sturm (2014), Physiscal properties of Arctic versus subarctic snow: Implications for high latitude passive microwave snow water equivalent retrievals, *J. Geophys. Res-Atmos.,* 119, 7254-7270.

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

Emms, P. W., and A.C. Fowler (1994), Compositional convection in the solidification of binary alloys, *J. Fluid Mech.,* 262, 111-139.

Fetterer, F., K. Knowles, W. Meier, and M. Savoie (2002), *Sea ice Index, updated daily.* National Snow and Ice data Center, Boulder, USA.

Flato, G. M., and W. D. Hibler III (1995), Ridging and strenght in modeling the thickness distribution of Arctic sea ice, *J. Geophys. Res.*, 100(C9), 18611-18626.

Francis, J. A., W. Chan, D. J. Leathers, J. R. Miller, and D. E. Veron (2009), Winter Northern Hemisphere weather patterns remember summer Arctic sea-ice extent, *J. Geophys. Res.*, 36, L07503.

Gent, P.R., G. Danabasoglu, L.J. Donner, M.M. Holland, E.C. Hunke, S.R. Jayne, D.M. Lawrence, R.B. Neale, P.J. Rasch, M. Vertenstein, P.H. Worley, Z. Yang, and M. Zhang (2011), The Community Climate System Model veersion 4, *J. Climate*, 24,4973- 4991.

Golden, K. M., H. Eicken, A. L. Heaton, J. Miner, D. J. Pringle, and J. Zhu, J. (2007), Thermal evolution of permeability and microstructure in sea ice, *Geophys. Res. Lett.,* 34, L16501.

Griffies, S. M., A. Biastoch, C. Boning, C., F. Bryan, G. Danabasoglu, E. P. Chassignet, M. H. England, R. Gerdes, H. Haak, R. W. Hallberg, W. Hazeleger, J. Jungclaus, W. G. Large, G. Madec, A. Pirani, B. L. Samuels, M. Scheinert, A. S. Gupta, C. A. Severijns, H. L. Simmons, A. M. Tregier, A. M. Winton, S. Yeager, and J. Yin (2009), Coordinated ocean-ice reference experiments (COREs), *Ocean Mod.,* 26, 1-46.

Halton, J. H. (1960), On the effciency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math.,* 2, 84-90.

Hastie, T., and R. Tibshirani (1986), Generalized Additive Models, *Stat. Sci.,* 1(3), 297-318.

Hewitt, H.T., D. Copsey, I.D. Culverwell, C.M. Harris, R.S.R. Hill, A.B. Keen, A.J. McLaren, and E.C. Hunke (2011), Design and implementation of the infrastructure of HadGEM3: The next generation Met Office climate modeling system, *Geosci. Model Dev.*, 223-253, doi:10.5194/gmd-4-223-2011.

Holland, M.M., M.C. Serreze, and J. Stroeve (2008), The sea ice mass budget of the Arctic and its future change as simulated by coupled climate models, *Clim. Dyn.*, 34, 185-200.

Holland, M., D. A. Bailey, B. P. Briegleb, B. Light, and E. C. Hunke (2012), Improved sea ice shortwave radiation physics in CCSM4: The impact of melt ponds and aerosols on Arctic sea ice, *J. Climate,* 25, 1413-1430.

Homma, T., and A. Saltelli (1996), Importance measures in global sensitivity analysis of nonlinear models, *Reliab. Eng. Syst. Safe.,* 52, 1-17.

Hopkins, M. A. (1998), Four stages of pressure ridging, *J. Geophys. Res.,* 103, 21883-21891.

Hora, S. C., and J. C. Helton (2003), A distribution-free test for the relationship between model input and output when using Latin hypercube sampling, *Reliab. Eng. Syst. Safe.,* 79, 333-339.

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

Hunke, E. C., and M. Holland (2007), Global atmospheric forcing data for Arctic ice-ocean modeling, *J. Geophys. Res.,* 112, C06S14.

Hunke, E. (2010), Thickness sensitivities in the CICE sea ice model. *Ocean Model.,* 34, 137-149.

Hunke, E. C., D. A. Hebert, and O. Lecompte (2013), Level-ice melt ponds in the Los Alamos sea ice model, CICE, *Ocean Mod.,* 71, 26-42.

Hunke, E. C., W. H. Liscomb, A. K. Turner, N. Jeffery, and S. Elliot (2015), *CICE: the Los Alamos Sea Ice Model, Documentation and Software User's Manual, version 5.1.* Tech. Rep. LA-CC-06-012, Los Alamos, New Mexico.

Jacques, J., C. Lavergne, and N. Devictor (2006), Sensitivity analysis in presence of model uncertainty and correlated inputs, *Reliab. Eng. Syst. Safe.,* 91, 1126-1134.

Jones, D. W. R., and M. G. Worster (2014), A physically based parameterization of gravity drainage for sea-ice modeling, *J. Geophys. Res-Oceans,* 119, 5599-5621.

Kennedy, M. C., and A. O'Hagan (2001), Bayesian calibration of computer models, *J. R. Stat.*

Soc., 63, 425-464.

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

Koch, C. (1988), A coupled sea ice-atmospheric boundary layer model, part I: description of the model and 1979 standard run, *Beitr. Phys. Atmos.,* 61, 344-354.

Krembs, C., H. Eicken, and J. W. Deming (2011), Exopolymer alteration of physical properties of sea ice and implications for ice habitability and biogeochemistry in a warmer Arctic, *PNAS,* 108(9), 3653-3658.

Kurtz, N. T., and T. Markus (2012), Satellite observations of Antarctic sea ice thickness and volume, *J. Geophys. Res.,* 117, C08025.

Langehaug, H.R., F. geyer, L.H. Smedsrud, and Y. Gao (2013), Arctic sea ice decline and ice export in the CMIP5 historical simulations, *Ocean Mod.,* 71, 114-126.

Large, W. G., and S. G. Yeager (2009), The global climatology of an interannually varying air-sea flux data set, *Clim. Dyn.,* 33, 341-364.

Lecompte, O., T. Fichefet, M. Vancoppenolle, F. Domine, and F. Massonnet (2013), On the formulation of snow thermal conductivity in large-scale sea ice models, *J. Adv. Model. Earth Syst.,* 5, 542-557.

Lemieux, C. (2009), *Monte Carlo and Quasi-Monte Carlo sampling*, Springer, New York.

Lepparanta, M. (2011), *The drift of sea ice*, Springer-Verlag, Berlin.

Light, B., D. Perovich, M. Webster, C. Polashenski, and R. Dadic (2015), Optical properties of melting first-year Arctic sea ice, *J. Geophys. Res. Oceans*, 120, 7657-7675, doi:10.1002/2015JC011163.

Liscomb, W. H., E. C. Hunke, W. Maslowski, and J. Jakacki (2007), Improving ridging schemes for high-resolution sea ice models, *J. Geophys. Res.,* 112, C03S91.

Lu , P., Z. Li, B. Cheng, and M. Lepparanta (2011), A parameterization of the ice-ocean drag coefficient, *J. Geophys. Res.,* 116, C07019.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics,* 21(2), 239-245.

Meinander, O., S. Kazadzis, A. Arola, A. Riihelä, P. Räisänen, R. Kivi, A. Kontu, R. Kouznetsov, M. Sofiev, J. Svensson, H. Suokanerva, V. Aaltonen, T. Manninen, J-L.

Roujean, and O. Hautecoeur (2013), Spectral albedo of seasonal sow during intensive melt period at Sodankyla, beyond the Arctic Circle, *Atmos. Chem. Phys.,* 13, 3793-3810.

Nicolaus, M., C. Katlein, J. Maslanik, and S. Hendricks (2012), Changes in Arctic sea ice result in increasing light transmittance and absortion, *Geophys. Res. Lett,.* 39, L24501.

Nield, D. A., and A. Bejan (2006), *Convection in porous media*, Springer, New York, USA.

Notz, D., and M. G. (2009), Desalination processes of sea ice revisited, *J. Geophys. Res.,* 114, C05006.

O'Hagan, A. (2006), Bayesian analysis of computer code outputs: A tutorial, *Reliab. Eng. Syst. Safe.,* 91, 1290-1300.

O'Hagan, A., and M. West (2013), *The Oxford Handbook of Applied Bayesian Analysis*. Oxford University Press, New York.

Oakley, J. E., and A. O'Hagan (2004), Probabilistic sensitivity analysis of complex models: a Bayesian approach, *J. R. Statist. Soc. B,* 66(3), 751-769.

Owen, A. (1998), Scrambling Sobol' and Niederreiter-Xing points, *J. Complexity,* 14, 466-489.

Owen, A. B. (2013), Better estimation of small Sobol' sensitivity indices. *ACM T. Model. Comput. S.,* 23(2), 11:1-11:17.

Perovich, D. K., T. C. Grenfell, J. A. Richter-Menge, B. Light, W. B. Tucker III, and H. Eicken (2003), Thin and thinner: Sea ice mass balance measurements during SHEBA, *J. Geophys. Res.,* 108(C3), 8050.

Peterson, K., P. Bochev, and B. Paskaleva (2010), Development, sensitivity analysis, and uncertainty quantification of high-fidelity Arctic sea ice models, Report SAND2010-6218, Sandia National Laboratories.

Rae, J.G.L., H.T. Hewitt, A.B. Keen, J.K. Ridley, J.M. Edwards, C.M. Harris (2014), A sensitivity study of the sea ice simulation in the global coupled climate model, HadGEM3, *Ocean Mod.*, 74, 60-76.

Rasmussen, C. E., and C. K. I. Williams (2006), *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, USA.

Röske, F. (2001), An atlas of surface fluxes based on the ECMWF reanalysis. A climatological data set to force global ocean general circulation models, *Tech. Rep. 323*, Max-Planck-Inst. Für Meteorol., Hamburg, Germany.

Rothrock, D. A. (1975), The energetics of the plastic deformation of pack ice by ridging, *J.*

Geophys. Res., 80, 4514-4519.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004), *Sensitivity Analysis in Practice: A guide to assessing scientific models*, John Wiley & Sons Ltd., West Sussex, UK.

Saltelli, A., and P. Annoni (2010), How to avoid a perfunctory sensitivity analysis, *Environ. Modell. Softw.,* 25, 1508-1517.

Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola (2010), Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index, *Comput. Phys. Commun.,* 181, 250-270.

Screen, J. A., and I. Simmonds (2010), The central role of diminishing sea ice in recent Arctic temperature amplification, *Nature*, 464, 1334-1337.

Seuront, L., S. C. Leterme, J. R. Seymour, J. G. Mitchell, D. Ashcroft, W. Noble, P. G. Thomson, A. T. Davidson, R. van der Enden, F. J. Scott, S. W. Wright, M. Schapira, C. Chapperon, and N. Cribb (2010), Role of microbial and phytoplanktonic communities in the control of seawater viscosity off East Antarctica (30º-80º E), *Deep-Sea Res. II,* 57, 877-886.

Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simulat.,* 55, 271-280.

Stössel, A. (1992), Sensitivity of Southern Ocean sea-ice simulations to different atmospheric forcing algorithms, *Tellus,* 44A, 395-413.

Sturm, M., J. Holmgren, M. König, and K. Morris (1997), The thermal conductivity of seasonal snow, *J. Glaciol.,* 43(143), 26-41.

Taylor, P.C., M. Cai, A. Hu, G.A. Meehl, W. Washington, and G.J. Zhang (2013), A decomposition of feedback contributions to polar warming amplification, *J. Climate*, 26, 7023-7043.

The GPy authors (2012-2015), *GPy: A gaussian process framework in python.*

Thomas, D. N., and G. S. Dieckmann (2009), *Sea Ice*, John Wiley & Sons, Ltd., Chichester, UK.

Thorndike, A. S., D. A. Rothrock, G. A. Maykut, and R. Colony (1975), The thickness distribution of sea ice, *J. Geophy. Res.*, 80, 4501-4513.

Tonani, M., M. Balmaseda, L. Bertino, E. Blockley, G. Brassington, F. Davidson, Y. Drillet, P. Hogan, T. Kuragano, T. Lee, A. Mehra, F. Paranathara, C.A.S. TAnajura, and H. Wang (2015), Status and future global and regional ocean prediction systems, *J. Oper. Oceanogr.,* 8, S2, s201-s220.

Tsamados, M., D. L. Feltham, D. Schroeder, D. Flocco, S. L. Farrel, N. Kurtz, S. W. Laxon, and S. Bacon (2014), Impact of variable atmospheric and oceanic form drag on simulations of Arctic sea ice, *J. Phys. Oceanogr.,* 44, 1329-1353.

Turner, A. K., E. C. Hunke, and C. M. Bitz (2013), Two modes of sea-ice gravity drainage: A parameterization for large-scale modeling, *J. Geophys. Res.,* 118, 2279-2294.

Uotila, P., S. O'Farrell, S. J. Marsland, and D. Bi (2012), A sea-ice sensitivity study with a global ocean-ice model, *Ocean Model.,* 51, 1-18.

Vage, K., R. S. Pickart, V. Thierry, G. Reverdin, C. M. Lee, B. Petrie, T. A. Agnew, A. Wong, and M. H. Ribergaard (2008), Surprising return of deep convection to the subpolar North Atlantic Ocean in winter 2007-2008, *Nat. Geosci*., 2, 67-72.

Vancoppenolle, M., S. Bouillon, T. Fichefet, H. Goosse, O. Lecompte, M. A. Morales Maqueda, and G. Madec (2012), *LIM TheLouvain-la-Neuve Sea Ice Model.*, Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL).

Vickers, D., and L. Mahrt (2006), Evaluation of the air-sea bulk formula and sea-surface temperature variability from observations, *J. Geophys. Res.,* 111, C05002.

Weeks, W. F., and W. D. Hibler (2010), *On sea ice*, University of Alaska Press, Fairbanks.

Wiebe, H., G. Heygster, E. Zege, T. Aoki, and M. Hori (2013), Snow grain size retrieval SGSP from optical satellite data: Validation with ground measurements and detection of snow fall events, *Remote Sens. Environ.,* 128, 11-20.

Wiscombe, W. J., and S. G. Warren (1980), A model for spectral albedo of snow. I: Pure snow, *J. Atmos. Sci,.* 37, 2712-2733.

Wood, S. N. (2000), Modelling and smoothing parameter estimation with multiple quadratic penalties, *J. R. Statist. Soc. B,* 62(2), 413-428.

Wood, S. (2006), *Generalized Additive Models: An Introduction with R*, Chapman & Hall/CRC, Boca Raton, USA.

Wood, S. (2015), *Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation*.

Zygmuntowska, M., P. Rampal, N. Ivanova, and L. H. Smedsrud (2014), Uncertainties in Arctic sea ice thickness and volume: new estimates and implications for trends, *Cryosphere,* 8, 705-720.

Tables

Table 1. Uncertain input parameters in CICE v5.1 for the configuration used in this study. The parameters have been grouped in broad categories indicating the model physics they more directly affect.

rhos	snow density
rhoi	ice density

Melt ponds	
rfracmin	min. fraction of melt water added to ponds
rfracmax	max. fraction of melt water added to ponds
pndaspect	aspect ratio of pond changes
hs1	snow depth transition to pond ice

Ridging	
astar	participation function e-folding scale
mu_rdg	redistribution parameter
C_s	energy dissipated due to shear
C_f	frictional dissipation
fsnowrdg	snow fraction surviving ridging

Dynamics	
iceruf	roughness for neutral air-ice drag
dragio	neutral ocean-ice drag
ecc	yield curve aspect ratio

Radiation	
rsnw_fresh	fresh snow grain size
rsnw_mlt	max. melting snow grain size
rsnw_nonmelt	non-melting snow grain size
rsnw_sig	sigma of snow grain size
dT_mlt	melt/no-melt snow grain Δtemp.
hi_ssl	ice scattering layer
hs_ssl	snow scattering layer
hp0	pond depth for transition to bare ice.
R_ice	sigma coeff. for albedo of bare ice
R_pnd	sigma coeff. ponded ice albedo.
kalg	algae absorption coefficient.
R_snw	sigma coeff. for snow grain
fr_min	overcast factor for snow grain

Mushy thermodynamics	
kb	thermal conductivity of brine
kappal	thermal diffusivity of brine
ksno	thermal conductivity of snow
cp_ice	sea ice heat capacity
phi_i_mushy	solid fraction at lower interface
a_rapid_mode	brine channel diameter
Rac_rapid_mode	critical Rayleigh number
dSdt_slow_mode	slow drainage strength
phic	liquid fraction for impermeability
advection_limit	max. fraction of brine advection
lambda_pond	drainage time-scale of ponds
viscosity_dyn	brine dynamic viscosity

Table 2. Prior probability distributions and range of variation of 39 CICE input parameters explored in the sensitivity analysis. The assumed probability density functions are indicated as N (normal), U (uniform), T (triangular), LN (log-normal), LU (log-uniform), LT (log-triangular), and LogitN (Logit normal).

name	**pdf**	**min**	**max**	**units**	**References**
rhos	N	260	410	kg m^{-3}	[*Derksen et al., 2014*]
rhoi	N	839	959	kg m^{-3}	[*Alexandrov et al., 2010*]
kb	U	0.49	0.6	W s^{-1} K^{-1}	[*Thomas and Dieckmann, 2009*]
kappal	U	3×10^{-8}	19×10^{-8}	m^2 s^{-1}	[*Notz and Worster, 2009*; *Emms and Fowler, 1994*]
ksno	U	0.03	0.65	W m^{-1} K^{-1}	[*Lecomte et al., 2013*; *Sturm et al., 1997*]
cp_ice	U	1800	3800	J kg^{-1} K^{-1}	[*Thomas and Dieckmann, 2009*]
phi_i_mushy	U	0.55	0.95	-	[*Turner et al., 2013*]
a_rapid_mode	LU	6×10^{-5}	4×10^{-3}	m	[*Weeks and Hibler, 2010*; *Turner et al., 2013*]
Rac_rapid_mode	U	1	50	-	[*Nield and Bejan, 2006*; *Jones and Worster, 2014; Turner et al., 2013*]
dSdt_slow_mode	LU	-1.7×10^{-5}	-1.7×10^{-10}	m s^{-1} K^{-1}	[*Turner et al., 2013*]
phi_c_slow_mode	LT	0.027	0.1	-	[*Gold et al., 2007; Turner et al., 2013*]
advection_limit	U	0.001	0.015	-	[*Hunke et al., 2015*]
lambda_pond	LU	1.15×10^{-8}	1.15×10^{-4}	s^{-1}	[*Hunke et al., 2015*]
viscosity_dyn	N	1.6×10^{-3}	3.6×10^{-3}	Kg m^{-1} s^{-1}	[*Krembs et al., 2011*; *Seuront et al., 2010*]
rfracmin	LogitN	0	1	-	[*Holland et al., 2012*] *Hunke et al., 2013*]
rfracmax	LogitN	0	1	-	[*Hunke et al., 2013*]
pndaspect	U	0.4	1.2	-	[*Perovich et al., 2003*; *Holland et al., 2012; Hunke et al., 2013*]
hs1	T	0	0.1	m	[*Wiscombe and Warren, 1980*; *Brandt et al., 2005*]
rsnw_fresh	T	20	300	μm	[*Aoki et al., 2000*; *Wiebe et al., 2013*]
rsnw_mlt	T	250	3000	μm	[*Meinander et al., 2013*]
rsnw_nonmelt	U	20	500	μm	[*Hunke et al., 2015*]
rsnw_sig	U	100	400	μm	[*Hunke et al., 2015*]

name	pdf	min	max	units	References
dT_mlt	U	1	2	K	[*Briegleb and Light, 2007*]
hi_ssl	LN	0.003	0.1	m	[*Briegleb and Light, 2007*; *Nicolaus et al., 2012*]
hs_ssl	LN	0.003	0.1	m	[*Briegleb and Light, 2007*]
hp0	U	0.0	0.30	m	[*Hunke et al., 2015*]
R_ice	U	-2	2	-	[*Briegleb and Light, 2007*]
R_pnd	U	-2	2	-	[*Briegleb and Light, 2007*]
kalg	LU	0.001	2.4	m^{-1}	[*Briegleb and Light, 2007*]
R_snw	U	-2	2	-	[*Briegleb and Light, 2007*]
fr_min	U	0.5	1.0	-	[*Briegleb and Light, 2007*; *Hunke et al., 2015*]
iceruf	LN	0.1×10^{-4}	19.8×10^{-4}	m	[*Koch, 1988*; *Stössel, 1992; Vickers and Mahrt, 2006; Andreas et al., 2010a; Andreas et al., 2010b*]
dragio	LN	0.2×10^{-3}	160×10^{-3}	-	[*Lu et al., 2011*]
ecc	T	0.5	5	-	[*Flato and Hibler III, 1995*; *Lepparanta, 2011*]
astar	U	0.01	0.07	-	[*Liscomb et al., 2007*]
mu_rdg	U	1	7	$m^{0.5}$	[*Liscomb et al., 2007*; *Hunke, 2010*; *Kim et al., 2006; Uotila et al., 2012*]
C_s	LogitN	0	1	-	[*Flato and Hibler III, 1995*]
C_f	U	0	50	-	[*Flato and Hibler III, 1995*; *Hopkins, 1998*]
fsnowrdg	LogitN	0	1	-	[*Vaconppenolle, 2012*; *Hunke et al., 2015*]

Figures

Figure 1. Frequency histograms of the 1980-2009 monthly mean sea ice extent (a, b), area (c, d), and volume (e, f) in the northern hemisphere during September (left) and March (right). Data includes model predictions from 302 model runs at different input parameter combinations according to the experimental design described in section 3.3.

Figure 2. Frequency histograms of the 1980-2009 monthly mean sea ice extent (a, b), area (c, d), and volume (e, f) in the southern hemisphere during September (left) and March (right), as in Figure 1.

Figure 3. The 1980-2009 monthly mean sea ice extent (a, b), area (c, d), and volume (e, f) in the ensemble of 302 model runs at different input parameter combinations according to the experimental design in section 3.3. The ensemble model mean +/- one standard deviation are given by blue dots and whiskers, respectively. For comparison, observational estimates are included as red dots [*Fetterer et al., 2002; Kurtz and Markus, 2012; Zygmuntowska et al., 2014*] and the estimates from a CICE run using default parameter values as yellow dots. The shaded region encloses the envelope of 302 ensemble model predictions.

Figure 4. Root-mean-square error (RMSE) of the Gaussian process emulators for each model output variable, fitted to monthly mean model predictions from 397 runs.

Figure 5. Comparison of mean emulation prediction and simulated September mean sea ice volume in the northern hemisphere in 9 cross-validation datasets. Diagonal lines indicate perfect agreement between emulated and simulated ice volume, and whiskers indicate one emulator standard deviation from the mean prediction.

Figure 6. Comparison of mean emulation prediction and simulated September mean sea ice volume in the southern hemisphere in 9 cross-validation datasets. Diagonal lines indicate perfect agreement between emulated and simulated ice volume, and whiskers indicate one emulator standard deviation from the mean prediction.

Figure 7. Seasonal variation of the total indices of six input parameters affecting uncertainty of sea ice extent, area, and volume in the northern hemisphere.

Figure 8. Seasonal variation of the total indices of four input parameters affecting uncertainty of sea ice extent, area, and volume in the southern hemisphere.

Figure 9. The first order (blue) and total (red) sensitivity indices of 39 input parameters affecting sea ice volume in the northern hemisphere during (a) September, (b) December, (c) March, and (d) June. The parameters are listed in descending order of importance for the end of summer (September) sea ice volume.

Figure 10. The first order (blue) and total (red) sensitivity indices of 39 input parameters affecting sea ice volume in the southern hemisphere during (a) September, (b) December, (c) March, and (d) June. The parameters are listed in descending order of importance for the end of summer (September) sea ice volume in the northern hemisphere as in Figure 9.

Figure 11. Main effects of the most important parameters affecting September sea ice volume in the northern hemisphere (Figure 9). The main effects (black curves) represent the first order terms in equation (1) as anomalies around an expected value (f_0). Red lines represent the 95% confidence intervals of the main effects. Horizontal axes are normalized for the minimum and maximum values for each parameter as in Table 2.

Figure 12. Second order interactions affecting September sea ice volume in the northern hemisphere including (a) *ksno-lambda_pond*, (b) *ksno-rsnw_mlt*, and (c) *ksno-R_snw*. The interactions represent second order terms in equation (1) as anomalies around an expected value (f_0). The figure axes are normalized between 0 and 1 for the minimum and maximum values for each parameter as in Table 2.