Bounded and categorized: targeting data assimilation for sea ice fractional coverage and non-negative quantities in a single column multi-category sea ice model

Molly M. Wieringa[1], Christopher Riedel[2], Jeffrey L. Anderson[2], and Cecilia M. Bitz[1]

[1]University of Washington, Department of Atmospheric Sciences, Seattle, WA, USA
[2]NSF National Center for Atmospheric Research, Data Assimilation Research Section, Boulder, CO, USA

Correspondence: Molly Wieringa (mmw906@uw.edu)

Abstract. A rigorous exploration of the sea ice data assimilation (DA) problem using a framework specifically developed for rapid, interpretable hypothesis testing is presented. In many applications, DA is implemented to constrain a modeled estimate of a state with observations. The sea ice DA application is complicated by the wide range of spatio-temporal scales over which key sea ice variables evolve, a variety of physical bounds on those variables, and the particular construction of modern complex sea ice models. By coupling a single-column sea ice model (Icepack) to the Data Assimilation Research Testbed (DART) ~~in a series of observing system simulation experiments (OSSEs)~~, the grid-cell response of a complex sea ice model ~~is explored with~~ to a range of ensemble Kalman DA methods designed to address the aforementioned complications is explored. The impact on the modeled ice-thickness distribution and the bounded nature of both state and prognostic variables in the sea ice model are of particular interest, as these problems are under-examined. Explicitly respecting boundedness has little effect in the winter months, but correctly accounts for the bounded nature of the observations, particularly in the summer months when prescribed SIC error is large. Assimilating observations representing each of the individual modeled sea ice thickness categories consistently improves the analyses across multiple diagnostic variables and sea ice mean states. These results elucidate many of the positive and negative results of previous sea ice data assimilation studies, highlight the many counter-intuitive aspects of this particular data assimilation application, and motivate better future sea ice analysis products.

1 Introduction

Recent rapid Arctic change has emphasized the influence of sea ice on the global climate system, our incomplete understanding of its recent history, and many shortcomings of current sea ice models. The tide of interest in addressing these issues is well-reflected in the accelerating application of data assimilation techniques in both sea ice reconstruction projects (Schweiger et al., 2011; Sakov et al., 2012; Mu et al., 2018; Williams et al., 2022) and modelling studies ~~(Zhang et al., 2021)~~ (Zhang et al., 2021; Korosov et al., 2023). Data assimilation, or DA, is a set of objective methods through which observations of a system are blended with a modeled estimate of that system. Through this blending, DA injects the ~~real-world~~ information gained via the observations, which are typically limited in space and can be intermittent in time, into a model capable of integrating that information forward in a spatio-temporally continuous, physically realistic manner. DA is most commonly used

to obtain accurate initial conditions for numerical weather prediction models, but can also be deployed in climate studies to reconstruct unobserved variables by synchronizing observable components of a system with nature (Brennan & Hakim, 2022) or to infer the correct parameterization values that should be used in earth system models (Zhang et al., 2021). To date, most sea ice DA applications have employed ensemble Kalman filtering (EnKF) methods, a family of DA algorithms based on the Kalman filter (Kalman, 1960; Evensen, 2003; Houtekamer & Zhang, 2016). EnKF methods approximate the application of a true Kalman filter by sampling the system of interest using model ensembles. In practical applications, the adjustments made by these filters can be considered in four steps~~.~~, which are outlined in Fig. 1 (adapted from Anderson, 2022) for a hypothetical adjustment of sea ice concentration (SIC). Firstly, the model is used to generate an ensemble of forecasts. Secondly, estimates of the observed quantities (e.g. SIC) are calculated from the model's state variables (e.g. categorized sea ice area fraction, $A_{ice,n}$). Thirdly, a version of the Kalman filter is applied to update the model's estimates of the observed quantity. Here, this will be referred to as observation-space incrementing. ~~Thirdly~~Finally, the adjustments made in observation space are used to determine the corresponding updates applied to the variables comprising the model state. This step is hereafter referred to as state-space regression. Together, observation-space incrementing and state-space regression are collectively known as filtering. ~~Lastly~~Once filtering is complete, the updated model state is then used to initialize the next forecast step. All together, this process is ~~termed~~ referred to as a DA cycle.

Substantial nuance can arise in the cycling process depending on the characteristics of the system in question. This makes DA in any earth system component model an intricate undertaking, and one often specifically tailored to the problem at hand. For sea ice this is particularly true, as sea ice models and observables unite many distinct challenges for DA in one system. ~~First~~Firstly, similar to atmospheric variables such as cloud fraction, sea ice variables tend to be bounded. For example, ice cannot be negatively thick; sea ice concentration (the fraction of a model grid cell covered by ice~~(sea ice concentration)~~) cannot fall below zero or exceed one. The Kalman methods applied to sea ice problems are based on assumptions that the model ensemble and the observation error distribution are normal distributions, which ~~thereby~~ linearizes the filtering process. For system variables that are bounded, however, the use of normal distributions in the filtering algorithm can produce adjustments during observation-space incrementing that violate physical bounds ~~.~~(as illustrated in step 3 of Fig. 1). When these violations are corrected (typically through a postprocessing step), the model ensemble mean is artificially shifted away from the bound, leading to ~~analysis inaccuracies~~a bias in the assimilation analysis. While non-Gaussian ensemble DA methods that avoid the use of normal distributions have been proposed, their application in high-dimensional systems has been limited (Riedel & Anderson, 2023; Anderson, 2010).
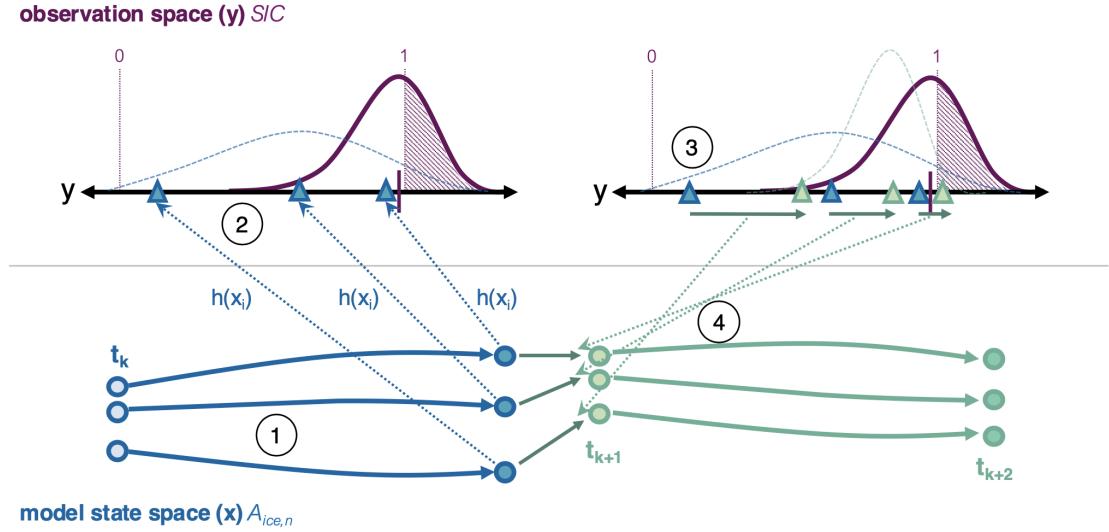
Figure 1. Schematic example of a data assimilation cycle in a sea ice application. A univariate example of a single data assimilation cycle is presented in four steps. In step 1, an ensemble of sea ice initial conditions (light blue circles) is forecasted forward from time t_k. In step 2 (dotted light blue arrows), the ensemble of sea ice states is translated into an ensemble of observational estimates (blue triangles) using a forward operator, $y = h(x)$. A continuous distribution shown by the light dashed blue line is fit to the observation estimates, and is compared to an observation, shown as a dark pink hash on the y-axis. The observational error distribution associated with that observation is shown in the dark pink curve, and regions of the error distribution that imply non-physical values are shaded with pink dots. In step 3, the model's observational estimates are adjusted by the data assimilation filter, based on the observation value and error distribution. The updated ensemble of observational estimates is shown in the green triangles. The amount that each ensemble member is adjusted is referred to as an observation increment (dark green arrows). The observation increments are then used to determine the state-space increments, or the amount that the model state variables need to be adjusted to be consistent with the observation increment; this step (represented here in the dotted light green arrows) is typically achieved by using the linear relationship between observations and state variables to project the observation increment into model state space. Once the model state has been adjusted, the model is reinitialized to produce the next forecast from t_{k+1} (light green circles to dark green circles). This figure has been reproduced with alterations from Anderson (2022).

Secondly, the relationship between variables observed in the real world and modeled in the sea ice state is not straightforward. Sea ice observing systems measure variables such as SIC or sea ice thickness (SIT). However, SIC and SIT are *diagnostic* in modern sea ice models, which typically evolve through an ice-thickness distribution (ITD). The ITD parameterizes sub-grid scale thermodynamic and mechanical processes that are strongly dependent on ice thickness (Bitz & Roe, 2004; Chevallier & Salas y Melia, 2012) by expressing the distribution of ice variables in a grid cell as functions of the ice thickness. In practice, the ITD describes a range of thicknesses within each grid cell and discretizes that range into an arbitrary number of thickness categories. Sea ice area and volume (and the snow volume atop the sea ice) are then similarly distributed across the thickness categories (Thorndike et al., 1975) and the evolutionary equations of the sea ice model are applied to each category individually. Observed SIC, SIT, and snow depth (SND) are aggregates of the "categorized" model variables of ice area ($A_{\mathrm{ice},n}$), ice volume ($V_{\mathrm{ice},n}$), and snow volume ($V_{\mathrm{sno},n}$), respectively; the latter three sets of variables represent the sea ice state. Thus, while estimates of SIC and SIT calculated in step 2 of the DA cycle are updated during observation-space incrementing when SIC or SIT observations are assimilated (step 3), the updates to the aggregate values are regressed out to each of the categorized variables during the state-space regression (step 4). The diagnostic SIC and SIT output at the end of the process are then re-aggregated from the updated categorized state variables; their accuracy relies not only on the direct filter updates on the aggregate quantities, but also on the model ensemble's relationship between the aggregated quantities and each of the categorized variables in the model state. Few studies have presented the impact of assimilating SIC or SIT on each of the model's categories individually, which raises the question of how well the process and impact of assimilating any observation into distribution-based sea ice models is understood. Recent work by Williams et al. (2022) documents the first attempt to assimilate an "observed" ice thickness distribution rather than just an aggregate observation into the sea ice component of a global climate model, with mixed results.

Both the non-Gaussian, bounded nature of sea ice and the relationship between aggregate observables and categorized state variables likely have important ramifications for sea ice DA but remain under-explored. This study presents a single-column sea ice data assimilation framework that allows for rapid hypothesis testing while retaining the thermodynamic physics and ITD of a complex sea ice model. Within this idealized framework, the impact of using DA algorithms that respect the boundedness of sea ice model variables and observations is explored, as is the ITD response of the model when assimilating aggregate versus categorized area and thickness observations. Section 2 provides an overview of the data assimilation framework and experimental methodology; Section 3 presents a discussion of the results generated by a suite of DA experiments targeting boundedness and categorized observations; Section 4 contextualizes this work with respect to more practical sea ice DA applications; Section 5 concludes.

2 Model and Methods

The data assimilation framework used in this study couples the Data Assimilation Research Testbed (DART, Anderson et al., 2009) to Icepack (version 1.3.1, Icepack, 2020), the column-physics package of the CICE sea ice model, which is widely used as the sea ice component of several Earth system models

85 ~~in~~ . Icepack can be run in a stand-alone ~~sea ice studies.~~ configuration as a sort of single-column model and is reviewed in Section 2.1. DART is discussed in more depth in Section 2.2. In keeping with naming conventions developed in coincident work (Riedel et al., 2023), ~~this framework~~ the collective assimilation system is referred to as CICE-SCM-DART. ~~A sea ice~~ All experiments performed for this study are observing system simulation experiments (OSSEs), which assimilate synthetic observations derived from a randomly-selected (and subsequently withheld) member of the sea ice ensemble. In

90 each experiment, the randomly selected member represents a known "true" state against which the efficacy of assimilating observations of various types and with various uncertainties can be evaluated. For simplicity, a sea ice quantity produced by ~~a~~ CICE-SCM-DART ~~experiment~~ is hereafter differentiated from the ~~corresponding synthetic observations assimilated~~ assimilated synthetic observations using the terms "modeled" and "observed", respectively.

2.1 Icepack

95 Icepack is maintained as the column physics module of CICE, with consistent thermodynamics, mechanical redistribution, and tracer support. ~~The model evolves conditions in four independent grid cells, each with a different surface type (land, open ocean, slab ice, or a categorized ITD). These four grid cells do not communicate in any way; as such, only output from the ITD grid cell is retained.~~

For use in the CICE-SCM-DART framework, 30 instances of Icepack are forced by unique atmospheric conditions extracted
100 from randomly selected members of ~~the CAM6 + DART~~ a recent large-ensemble reanalysis product (Raeder et al., 2021). ~~The sea ice model is tuned to the atmospheric forcing by setting the~~ Each instance of Icepack uses the mushy thermodynamics scheme (kitd = 1) and linear ITD remapping options (ktherm = 2), as well a Delta-Eddington shortwave radiative transfer scheme and the empirical CESM melt pond scheme. Dynamical forcing to the column is provided by sea ice deformation rates obtained from the SHEBA field campaign (Lindsay , 2002). The number of categories used in the ITD is set to 5. The snow
105 grain radius parameter (R_snw) is set to a value of -2. This choice ~~prevents discontinuous behavior in ice concentration related to ice-albedo feedback~~ , which is among the default values of (R_snw) used when CICE is coupled to an atmospheric model, avoids rapid refreezing events during the melt season ~~and thus allows for a wider summertime ensemble spread. The number of categories used in the ITD is set to 5.~~ that lead to unreasonably high summertime sea ice concentrations given the atmospheric forcing conditions. All other sea ice model parameters are held at their default values. Each instance of Icepack is also coupled
110 to a slab ocean; the ocean initial conditions and heat flux convergence forcing are ~~consistent across the~~ identical for all 30 members and are derived from the ocean component output of a fully-coupled historical simulation from the Community Earth System Model (CESM2). Both the ocean and atmosphere data sets represent grid cells nearest 75.54^oN, 174.45^oE, a point that straddles the East Siberian and Chuckchi Seas and experiences seasonal sea-ice advance and retreat. The use of a seasonal location for this case study allowing us to evaluate the performance of data assimilation near the upper and lower bounds of
115 sea ice concentration.

The ensemble is spun up over a 10-year period during which the atmospheric conditions cycle continuously over the year ~~2011.~~ 2011, allowing the sea ice simulations to diverge in response to atmospheric variability. No assimilation occurs during this period. Once spin-up is complete, a final year-long ensemble simulation is produced as a control case for the assimilation
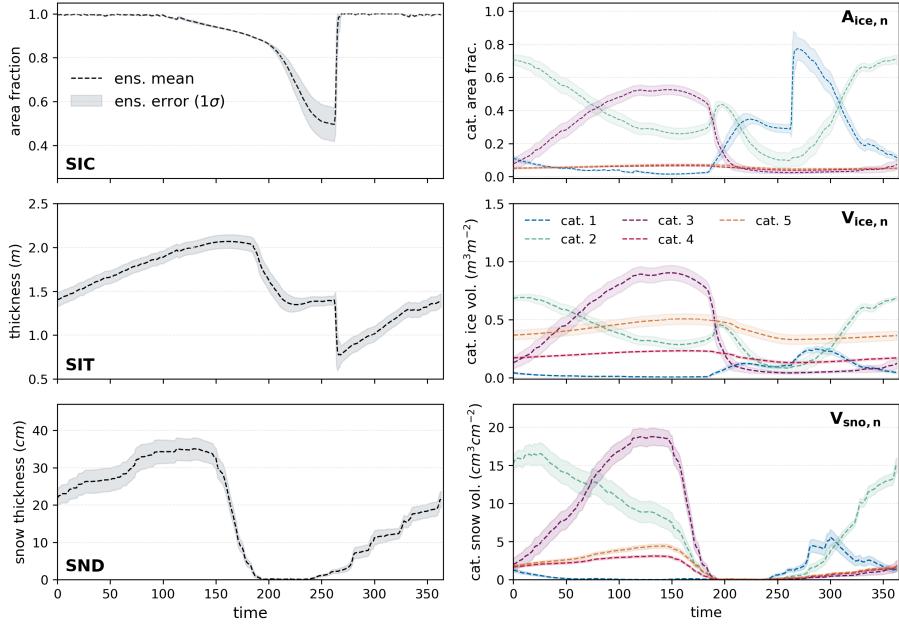
Figure 2. The FREE ensemble's three aggregate variables and the state variables from which they are derived. The aggregate variables shown on the left (SIC, SIT, and SND) are the area-weighted sums of the categorized state variables (category ice area, ice volume, and snow volume on the right. There are five thickness categories, where category 1 refers to the thinnest ice (0-0.64m) and category 5, the thickest ice (4.57m and thicker). Dark lines indicate the ensemble mean of each variable and lighter shading represents the ensemble standard deviation around the mean.

experiments. This simulation, which is also absent any assimilation, is hereafter referred to as the FREE case and is outlined in Fig. 2. Both categorized state variables (right) and their diagnosed aggregates (left) are shown, as both can be observed and adjusted by assimilation.

2.2 DART

DART is a modular data assimilation framework developed by the Data Assimilation Research Section at the NSF National Center for Atmospheric Research. DART interfaces with many models that range in complexity from the Lorenz 3-variable chaotic model to the Community Atmosphere Model (CAM6), the atmosphere component of the CESM2 climate model. DART implements the four-step cycling approach outlined in the introduction: forecast, conversion to observation space, observation-space incrementing, and state-space regression , and re-initialization. (Fig. 1). DART currently includes 10 filtering algorithms, including variants on the ensemble Kalman filter (EnKF; Evensen 2003) and several kernel and particle filter options. The default filter, the Ensemble Kalman Adjustment Filter (EAKF; Anderson 2001), implements a square-root filtering approach that increases the stability and efficiency of assimilating with smaller ensemble sizes compared to a traditional EnKF. Like most traditional ensemble filtering approaches, the EAKF makes Gaussian assumptions for the model ensemble and the observation error distributions.

Recently, Anderson (2022) developed a novel filtering approach known as the quantile-conserving ensemble filtering framework (QCEFF). QCEFF alters the process by which the updated ensemble is sampled from the analytical blend of the model ensemble distribution and the observation error distribution. As a result, DART users can prescribe non-Gaussian distributions that may better represent the model ensemble or observation of interest. For example, in the sea ice problem, QCEFF allows the user to prescribe distributions that respect sea ice bounds, a level of detail that cannot be attained by EAKF or other Gaussian filters. In this framework the user can prescribe a distribution for each observable or state variable, as well as differentiate the distribution used for observation-space incrementing versus state-space regression; this kind of choice allows the user to tailor the DA framework to the problem at hand in every step of the filtering process. When the user prescribes normal distributions in the QCEFF framework, the solution collapses to the EAKF.

We employ QCEFF to examine whether explicitly accounting for sea ice boundedness can improve sea ice assimilation analyses. To do so, we compare four different filtering approaches, outlined in Table 1. These filtering approaches use varying combinations of normal and piece-wise rank histogram distributions in the observation-space incrementing and state-space regression steps of the filter. Piece-wise rank histogram distributions prescribe no more information about the distribution of the sea ice system than can be gained from the discrete ensemble members themselves and can capture physical bounds; their use in DART's step 3 of the DART filtering algorithms and for sea ice applications is discussed in more detail in (Anderson, 2020), (Riedel & Anderson, 2023) and (Riedel et al., 2023)Anderson (2022), Riedel & Anderson (2023) and Riedel et al. (2023). The use of bounded normal rank histogram (BNRH) distributions in state-space regression (step 4) of the QCEFF enforces appropriate bounds by way of a series of transforms in probit and probability integral space. This aspect of the QCEFF also more deftly handles nonlinear relationships between observed quantities and modeled state variables and is addressed in (Anderson, 2023)depth for idealized cases in Anderson (2023).

2.3 Experimental Setup

All experiments performed for this study follow a perfect-model Observing System Simulation Experiment (OSSE) protocol (Zhang et al., 2018; Riedel & Anderson, 2023; Riedel et al., 2023), a methodology typically used to identify the impact of assimilating a set of proposed or synthetic observations. The use of synthetic observations allows for a close inspection of DA filter performance given a set of observations derived from a known state. Here, several different kinds of synthetic sea ice observations are assimilated using each of the filter types listed in Table 1. Each experiment was branched from the end of the ensemble spin-up period, assimilated observations for a year, and was then compared to the FREE case. The assimilation experiments presented in the results are listed in Table 2.

The synthetic observations assimilated (a subset of which are presented in Fig. 3) are identical across experiments and are derived from a randomly selected ensemble member of the FREE case, which is hereafter referred to as TRUTH. To capture the basic influence of observation instrument and algorithmic errors on sea ice DA, observation error magnitudes are estimated based on previous work (Zhang et al., 2018; Riedel et al., 2023) and expressed as a function of the daily TRUTH value (listed in Table 3). The error magnitude, which can be thought of as the second moment of a probability distribution, is then used to determine a prescribed observation error distribution (OED) centered on the TRUTH estimate of the observation. Each daily observation is then randomly sampled from the OED. The resulting observation timeseries thus captures reasonable noise around the known TRUTH. In ensemble Kalman DA studies preceding QCEFF, the OED was assumed to be a normal distribution around TRUTH values. Here, the OED is set as a bounded normal distribution, thereby accounting for the physical realities of sea ice observations.

Aggregate observation values extracted are SIT and SIC. The variance of the observation error distribution for each synthetic SIT observation is a linear function of the true SIT value on the order of tens of centimeters. Observation error variance for synthetic SIC observations is a parabolic function of the true value on the order of ten percent of grid cell area. As a result, observation error magnitudes when SIC declines in the summer months can be quite large, implying a plausible range of observations that may exceed the SIC upper bound of 1. When used to determine a bounded OED that does not exceed 1, these large errors lead to summer SIC observations that are biased low relative to TRUTH. The ramifications of this bias are discussed in Sections 3 and 3.1.

Synthetic categorized observations are also drawn from each of the model's area and volume ITD categories ($A_{\mathrm{ice},n}$ and $V_{\mathrm{ice},n}$) and are always assimilated together (i.e., assimilating $A_{\mathrm{ice},n}$ indicates that each of five area categories are assimilated simultaneously). Categorized area and volume observation error variances are assumed to follow a uniform distribution in each category, weighted by the total area (and midpoint thickness, in the case of volume observations) of that category. These errors are therefore generally less than 10% of the true category value (Fig. 3).

Because sea ice ensembles perturbed only by differing atmospheric conditions (and not by varying model parameters) are generally under-dispersive with respect to SIC (Zhang et al., 2018; Williams et al., 2022; Riedel & Anderson, 2023), we apply enhanced spatially-varying state-space prior inflation (El Gharamti et al., 2019) in each experiment. While the benefits of the spatial variation are lost on our application, the algorithm used implements an inverse gamma function that enables an increase
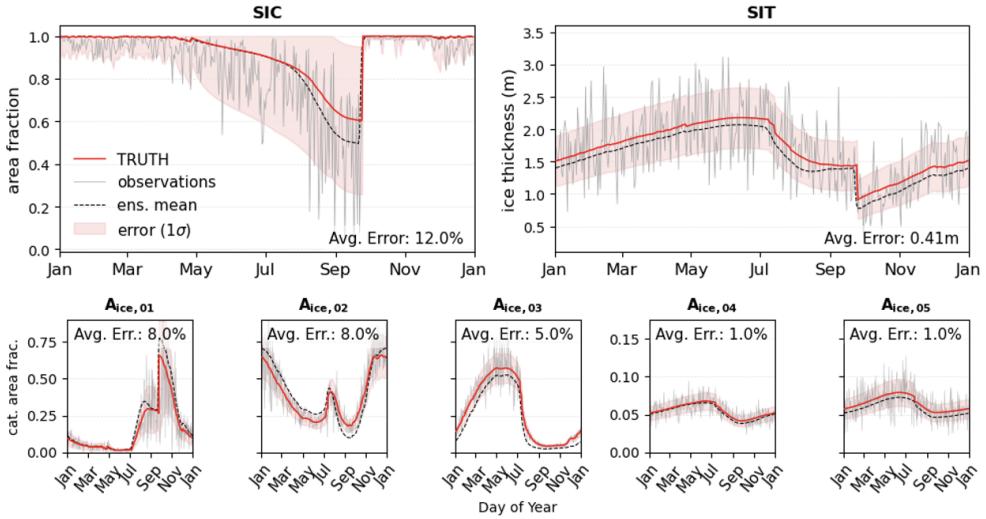
Figure 3. Synthetic observations extracted from a randomly selected member of the FREE ensemble. The observation assimilated are shown in grey lines for SIC (top left), SIT (top right) and category ice area ($A_{ice,01}$-$A_{ice,05}$) on the bottom row. The TRUTH from which the observations are generated in shown in the solid red line, while the FREE ensemble mean is shown in the dashed black line. The observation error standard deviation (1σ) is shown as red shading around TRUTH.

or decrease in ensemble spread and outperforms alternative inflation algorithms in some cases (El Gharamti et al., 2019). The applied inflation uses a damping factor of 0.9, a lower standard deviation bound of 0.6, and a maximum per-timestep standard deviation change of 5%.

190   Spatial localization is practically uninformative in a single-column application, but we explore the effect of "category localization" in the experiments assimilating $A_{ice,n}$ or $V_{ice,n}$. Category localization weights the covariance values between variables in different ITD categories by zero. As a result, an observation from any of the individual ITD categories is prevented from updating any state-space variable that is not also in ~~that~~ the same ITD category. In theory, this type of localization should limit the effects of potentially spurious relationships between categories and allow us to more reasonably treat category error

195   variances as uncorrelated.

Finally, since DA is not guaranteed to respect the physical bounds of a system, it is common to use some postprocessing method to correct any non-physical adjustments made by the filter. DART includes three postprocessing options for sea ice: two mass-aware re-scaling approaches and one rebalancing method that has been adapted from a CICE internal function (Riedel & Anderson (2023); the current default in CICE-SCM-DART). All experiments in Table 2 make use of ~~the CICE~~

200   ~~rebalancing option~~ this default rebalancing option, which redistributes the ice fractional coverage in each category to ensure that the thickness bounds are respected and then calculates consistent ice and snow volumes, salinities, and enthalpies once the updates have occurred. Each experiment was rerun using the other two postprocessing methods, but since no significant differences resulted, those additional experiments are not discussed here.

2.4 Evaluative Metrics

To evaluate results, the ensemble means of the FREE case and each experiment (EXP) in Table 2 are compared to TRUTH using three metrics: mean absolute error (MAE), root mean square error (RMSE), and the coefficient of efficiency (CE). The presented definitions are generalized such that EXP and TRUTH may represent the experiment ensemble mean and reference "true" value, respectively, of any of CICE-SCM's state or diagnostic variables. In this study, these metrics are applied only to to SIC, SIT, and SND.

MAE measures the average discrepancy between the forecast (FREE or EXP) and TRUTH over the course of the forecast period and is defined as

$$MAE = \sum_{i}^{n} \frac{|EXP_i - TRUTH_i|}{n}, \tag{1}$$

where n indicates the number of timesteps in the forecast period. RMSE, defined as

$$RMSE = \sqrt{\sum_{i}^{n} \frac{(EXP_i - TRUTH_i)^2}{n}}, \tag{2}$$

also evaluates how the forecast deviates from TRUTH but additionally provides a sense of whether the average discrepancy tends to include large outliers. RMSE is therefore always greater than MAE, but in a desirable forecast the difference between the two will be close to zero.

CE (Nash & Sutcliffe, 1970) measures forecast skill compared to TRUTH by evaluating how *efficient* the forecast is as a model of the observed system's mean and variance,

$$CE = 1 - \frac{\sum_{i}^{n} (EXP_i - TRUTH_i)^2}{\sigma_{TRUTH}^2}, \tag{3}$$

and lies between $-\infty$ and 1. CE equal to 1 indicates a perfect match between the forecast and the TRUTH (the numerator in the second term of Eq. 3 is zero), while CE of 0 reflects a forecast that performs only as well as climatological prediction (the deviations of the experiment from TRUTH are equal to the variance of the TRUTH around its mean). A negative CE indicates that the forecast is not skillful. In general, the more positive the CE value, the better the forecast.

To couch results in a generalized framework, differences in MAE and RMSE between the EXP forecasts and the FREE forecast are evaluated using a percent reduction approach, thereby diagnosing the impact of assimilating observations relative to forecast with no assimilation. For example, percent RMSE reduction (pRMSE) due to assimilating observations is calculated as

$$pRMSE = 100 \times \frac{RMSE_{FREE} - RMSE_{EXP}}{RMSE_{FREE}}. \tag{4}$$

230 ~~CE is a less intuitive metric from an error reduction perspective; the~~ Many of the experiments performed for this work have a high CE, due to the idealized nature of single-column OSSE experiments. In order to highlight the impact of assimilation~~in this metric is therefore quantified~~, we choose to quantify this metric as a CE increase (iCE),

$$iCE = CE_{EXP} - CE_{FREE}. \tag{5}$$

~~Statistically~~ In order to understand whether a) assimilating with different methods and different variables leads to meaningful
235 adjustments toward TRUTH and b) any combinations of observations and filters significantly outperform the others over the course of the year, statistically significant differences between ~~assimilation experiments~~the ensemble mean timeseries of each EXP, FREE, and TRUTH are diagnosed using a Welch's t-test.

3 Results

The results of assimilating observations of SIT, SIC, and categorized area $A_{\mathrm{ice},n}$ with an unbounded DA filter (f1_NORM) are
240 presented in Fig. 4. This case illustrates that CICE-SCM-DART replicates the results of larger modeling studies discussed in the Introduction. Assimilating SIT observations results in better sea ice analyses year-round than assimilating SIC observations, which have an impact only during the summer months when the model ensemble is capable of capturing variations in sea ice cover. In fact, assimilating SIC observations appears to have a negative impact on modeled SIC in Fig. 4, though this is because our method for producing synthetic SIC observations—which are derived using a bounded normal OED—generates
245 SIC observations that are biased low relative to the TRUTH (Fig. 3). This is particularly true in the summer months when modeled SIC in TRUTH is comparatively low and the prescribed observation error variance is large (Table ~~2~~3).

Unlike ~~in~~ the unbounded case ~~(Fig. 4)~~, when observations are assimilated with a fully bounded filter (f101_BNRH) the bounded ~~observation error distribution~~ OED is appropriately accounted for and the results lie close to the FREE mean (Fig. 5). From this we conclude that while a bounded filter does not overcome the limited efficacy of assimilating SIC observations,
250 respecting boundedness in the assimilation does prevent the introduction of additional bias related to assumptions about the OED. We also note that assimilating SIC observations with $1/10^{th}$ of the error prescribed in Table ~~2~~3 does shift the resulting modeled SIC closer to TRUTH (not shown), though whether such small magnitude errors are reasonable is a separate discussion left for other work. In contrast, assimilating $A_{\mathrm{ice},n}$ observations performs at least as well as assimilating SIT observations in the unbounded case, and will be discussed in more depth later.
255 A more succinct comparison of the experiments listed in Table 2 is presented in Fig. 6. In terms of modeled SIT, we find that the assimilation of any observation that either explicitly or implicitly (through categorization in the ice thickness distribution) contains information about ice thickness reduces MAE by between ~~80 and 95~~70 and 90% and improves the CE score by ~~∼0.3~~0.1, regardless of the filter used. ~~Assimilating categorized observations tends to outperform assimilating SIT in terms of pMAE, but falls slightly short according to iCE. This discrepancy across metrics implies that assimilating categorized~~
260 ~~observations may not capture as much of the observed variance in SIT as assimilating SIT observations, a conclusion supported by the relative pRMSE achieved in each case (Fig. A1).~~ Experiments assimilating SIT and categorized observations are not
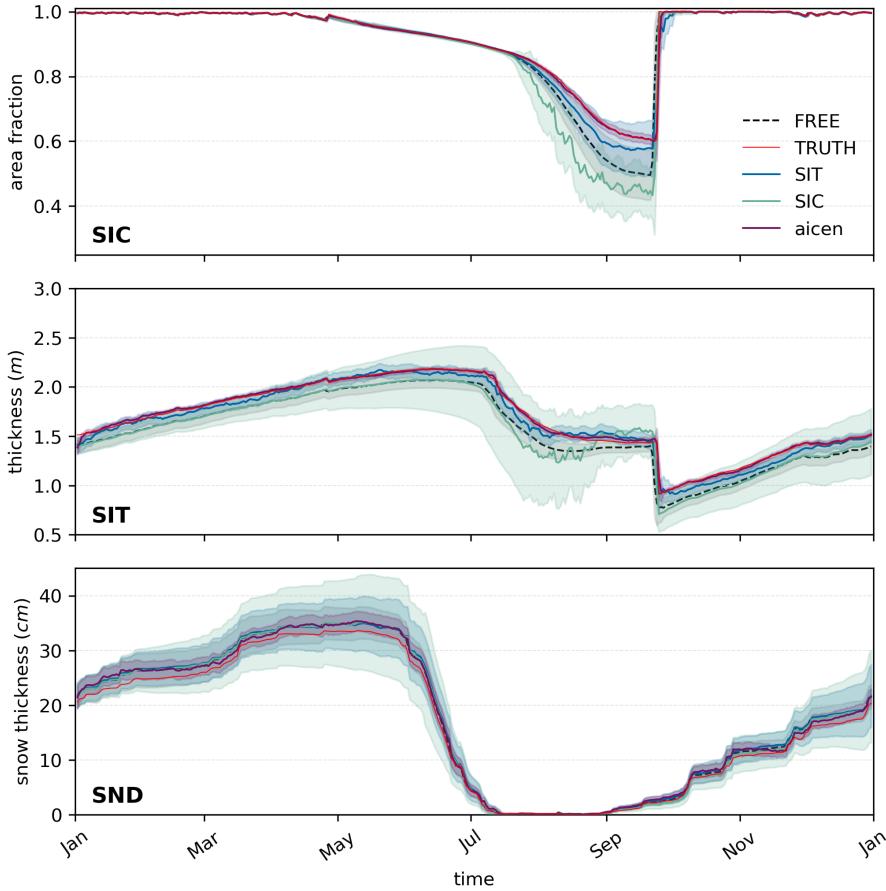
Figure 4. Assimilating with unbounded algorithms. The results of using an unbounded filter (f1_NORM) to assimilate SIC (solid pink line), SIT (solid teal) or category area observations (solid green) are shown for modeled SIC (top panel), SIT (middle panel) and SND (bottom panel). The black line represents the FREE case (without assimilation) and the thin red lines are the randomly selected TRUTH. For the results shown, thick lines are ensemble means and shading represents the ensemble standard deviation around the mean. Observations are assimilated at daily intervals throughout atmospheric forcing year 2011.

significantly different from ~~one another or~~ TRUTH, though they are all significantly different from the FREE ensemble mean (Fig. 7).

Adjustments to modeled SIC are more variable. The relative lack of improvement as a result of assimilating SIC compared to SIT is not a novel result (Blockley & Peterson, 2018; Kimmeritz et al., 2018; Mu et al., 2018; Zhang et al., 2018; Fiedler et al., 2022; Willi , but a good confirmation that the grid-cell level responses investigated here are reminiscent of sea ice DA studies that use more traditional ensemble filtering methods and assimilate on larger grids. For modeled SIT and SND, there is very little variation in the results as a function of the filter used (Fig. 6). For modeled SIC, larger pMAE tends to stem from cases using ~~unbounded regression methods (NORM) and when assimilating SIC observations~~ to update modeled SIC, using a bounded
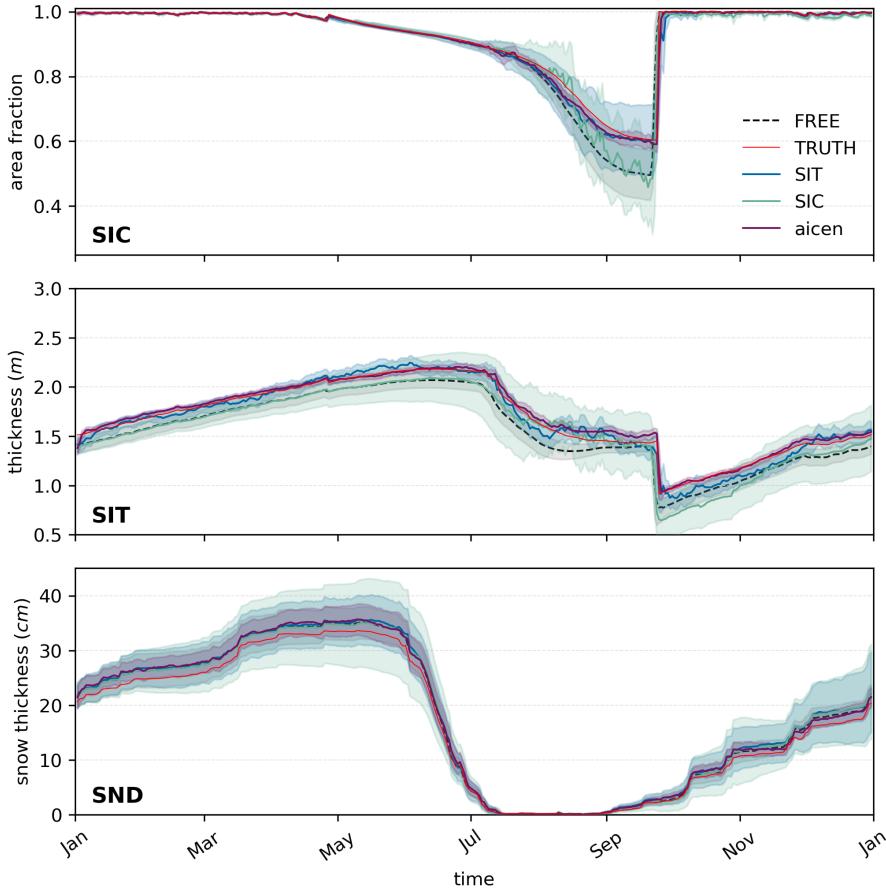
Figure 5. ~~Assimilating with unbounded algorithms.~~ Assimilating with bounded algorithms. Same as Fig. 4 but for case in which observations are assimilated using the f101_BNRH (fully bounded) filter.

270   ~~filter in observation space (f101) leads to notable improvements compared to using an unbounded filter (~~totally unbounded or totally bounded filtering (f1~~)~~_NORM or f101_BNRH) algorithms, or when assimilating categorized observations.

Finally, modeled SND is degraded by the assimilation of sea ice observations in all cases, except those which assimilated categorized observations with a totally bounded filter. Assimilating snow depth observations has been shown to improve snow estimates in large models when compared to cases in which snow was updated only via postprocessing (Riedel & Anderson, 275   2023), as well as in a single-column model when assimilated alongside sea ice observations (Riedel et al., 2023). In the experiments performed here, categorized snow (*vsnon*) is a state variable that is updated via regression with the model's observed quantities but no snow observations are assimilated. The general inefficacy of sea ice observations to reduce snow bias likely derives from an ensemble relationship between sea ice variables and categorized snow that produces too much late winter/early spring snow on thicker ice and too little on thinner ice (Fig. 8).
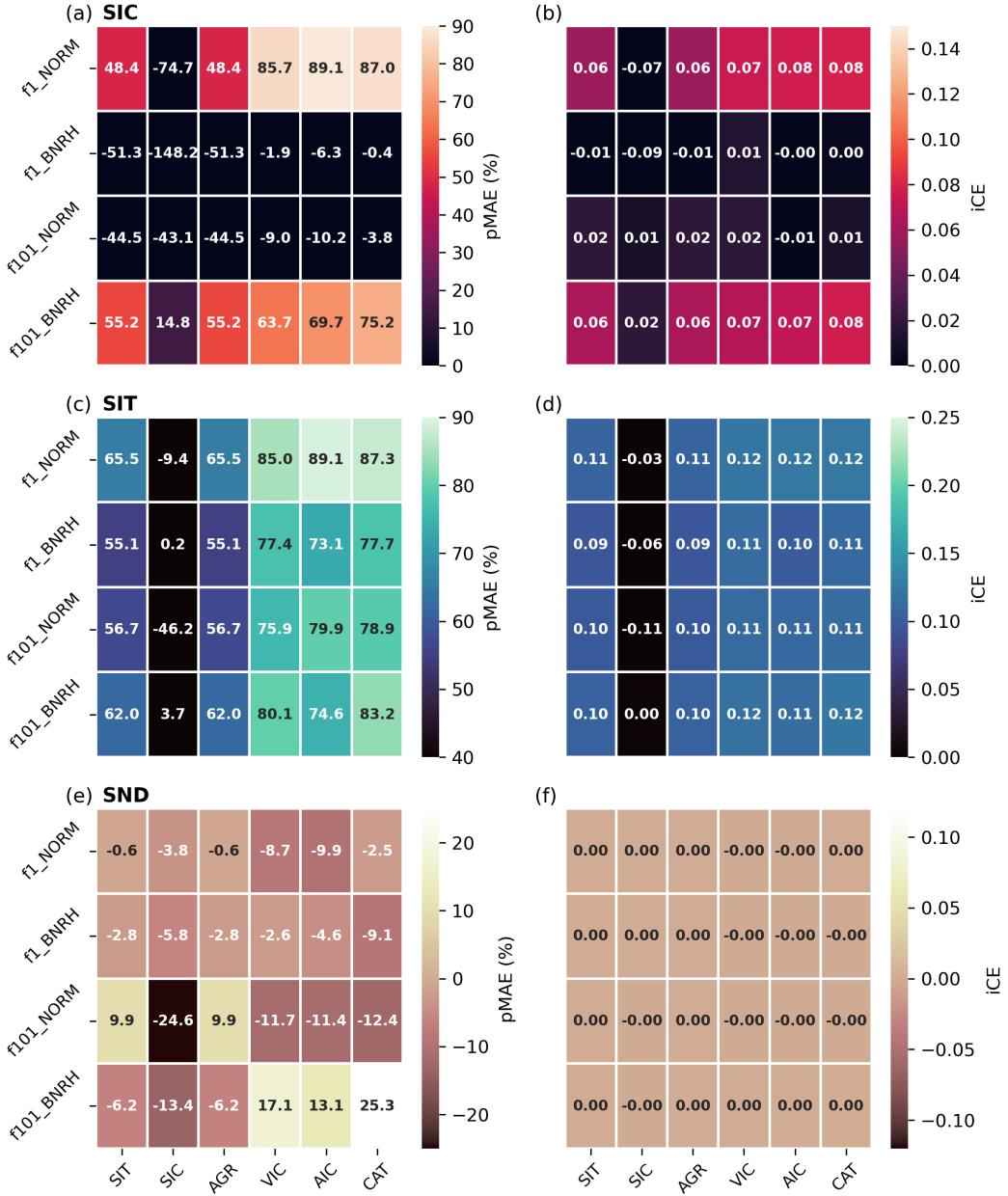
Figure 6. Bias reduction and model efficiency as a function of filter type and observation kind. Percent MAE reduction (pMAE) (left column) and CE increase (iCE) (right column) relative to the FREE forecast as a result of assimilating various observation kinds (x-axis, see Table 2 for definitions) with each filtering method (y-axis). Results are shown for modeled SIC (top row), SIT (middle row), and SND (bottom row). In general, lighter-toned colors indicate a more beneficial impact due to assimilation than darker-toned colors. The number values indicate the specific pMAE or iCE associated with each experiment.
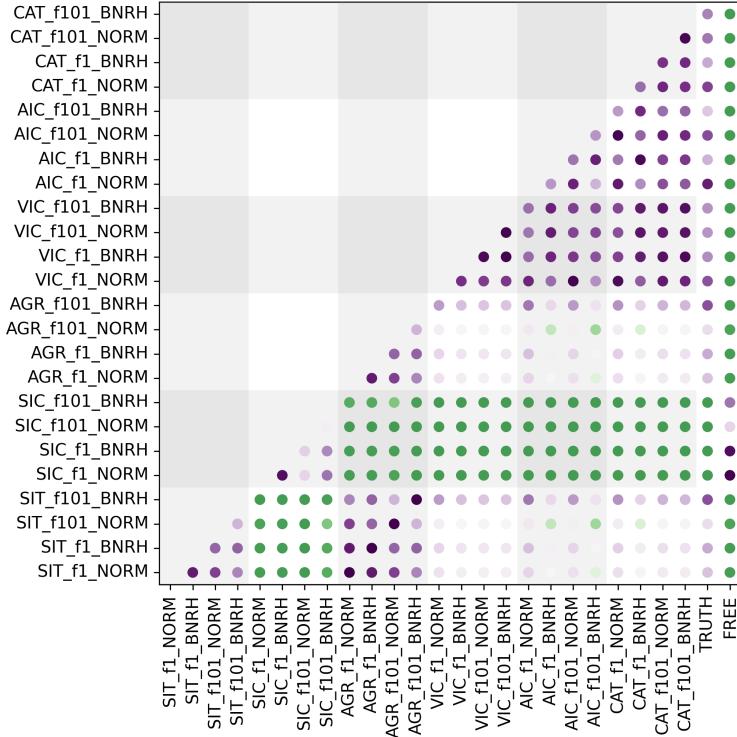
Figure 7. Significant differences between experiments. The significance of deviations between each assimilation experiment, TRUTH, and the unassimilated FREE ensemble mean are shown. The color gradient represents p-values for the statistical difference between each experiment shown on the x-axis with each experiment shown on the y-axis. The rightmost columns show the p-values for differences from TRUTH and the FREE case. Purple shades indicate insignificant difference at a p-value of 0.05, while greens indicate that the two cases in question differ significantly at a p-value of 0.05.

3.1 Boundedness

In general, we find the metrics in Figs. 6 and 7 have a rather weak dependence on whether or not the filter respects bounds for modeled SIT and SND, especially when compared to the obvious dependence on the kind of observation assimilated. There is essentially no dependency highlighted by iCE, and only minimal variation in pMAE. In terms of modeled SIC, however, the impact of using a bounded filter is more apparent (Fig. 9). The use of bounded rank histogram distributions in observation-space allows the filter to correctly infer the bounded nature of the observation error distribution (which respects the physical upper bound of 1 for SIC) and its relationship to TRUTH. The adjustments thus avoid degrading modeled SIC and lead to a positive annual pMAE (Fig. 6) and reduced bias relative to TRUTH, particularly in the melt season, when SIC observation errors are particularly large (Figs. 9, 5). The poor performance of the intermediary filters (f1_BNRH and f101_NORM) to constrain modeled SIC (Fig. 6) can be attributed to their inability to adjust SIC to total ice cover in the winter months (not shown).
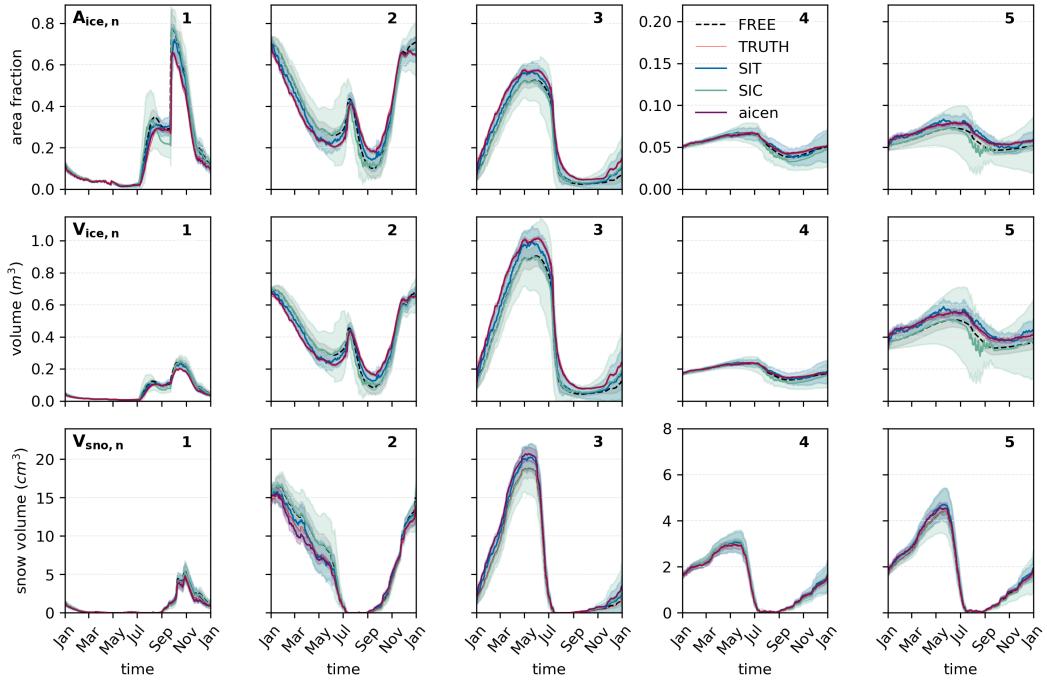
Figure 8. Category-level impact of assimilating with unbounded algorithms. Same as Fig. 4 but for each of the model's area (top row), volume (middle row) and snow volume categories (bottom row). TRUTH (the thin red line) may be difficult to identify in some panels, as the cases assimilating SIT and $A_{ice,n}$ (solid teal and green lines) lie very close to TRUTH.

The under-performance of ~~bounded filters away from the upper-bound of~~ the bounded filters with respect to SIC is likely due to the nature of the model state variables (categorized ice area, ice volume, and snow volume). Recall that the values being diagnosed (SIC, SIT, and SND) are calculated from categorized quantities using forward operators, but are not themselves state variables. This formulation leads to an issue with properly constraining modeled SIC. In the first step of the assimilation, bounds are placed on the observed quantity, SIC, which is calculated by applying a forward operator (a simple summation) to the model's forecast of the category area fractions in the ITD. Observation space incrementing respects the bounds prescribed on the observable. However, in the second step of the assimilation, the increment calculated between the observation and the model's estimate of the observed quantity is mapped back onto the category-based state variables using regression. This step also respects boundedness, but must rely on bounds prescribed by the user for each of the state variables. The only objective bounds that can be placed on each individual category area fraction are [0, 1], meaning that the regression of the observation-space increment can update each of the individual category area fractions to a value anywhere in that range. However, diagnostic SIC used to evaluate the forecast is calculated anew from the adjusted category area fractions, and is therefore no longer constrained on [0, 1], but rather on [0, 5]. As such, while the bounded filters respect the imposed bounds on both observed and state variables as intended (not shown), the dependency of the sea ice model on the prescribed ITD categories confounds an attempt to truly respect upper bounds on SIC.
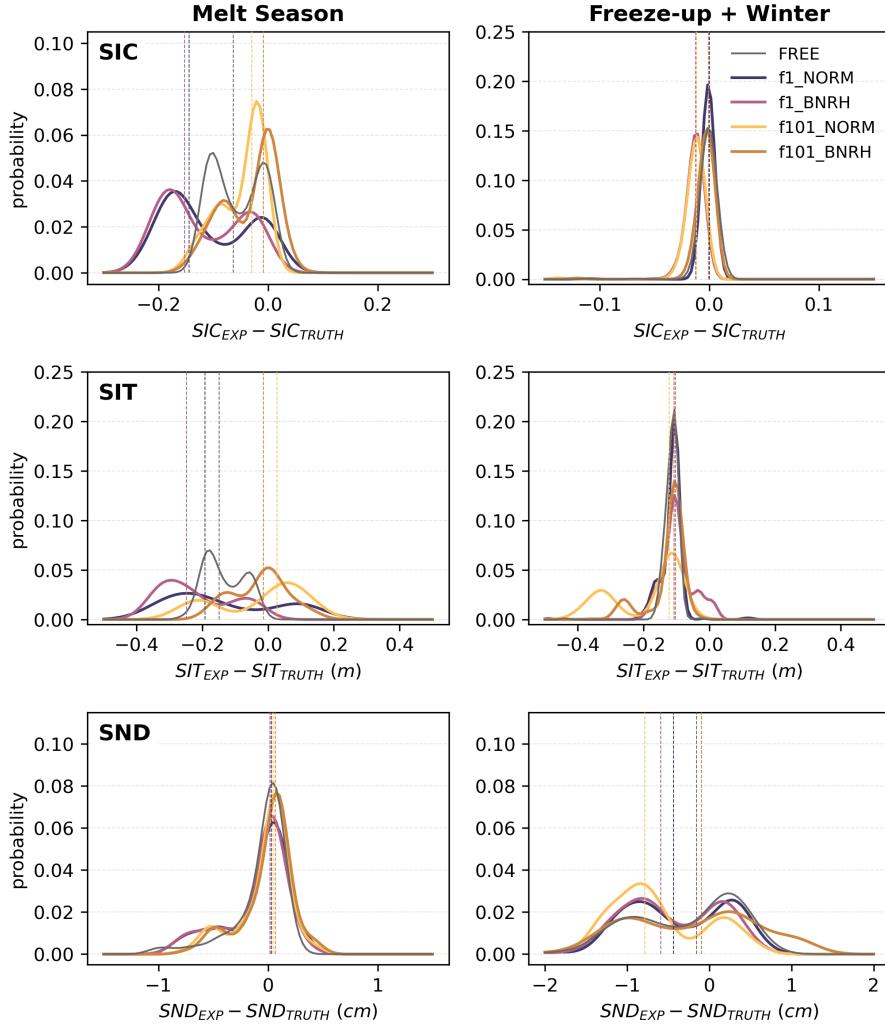
Figure 9. Seasonal impact of using bounded filters when assimilating SIC observations. The normalized probability density functions of the differences between the EXP mean and TRUTH in each assimilation cycle are shown for modeled SIC (top row), SIT (middle row) and SND (bottom row), along with their respective sample medians (dashed vertical lines). The dark grey distribution in each panel represents the difference between the FREE ensemble mean and TRUTH as a reference. The differences are divided into the melt season (left column, July 1st - September 15th) and the rest of the year (right column) and highlight the positive seasonal impact of using bounded algorithms—in the melt season, the distributions are shifted closer to a EXP-TRUTH difference of zero when a bounded filter is used (yellow lines) than when an unbounded one is used (pink lines). This effect is most prominent in the melt season months because the uncertainties associated with the assimilated SIC observations are largest in these months, and thus the bounded synthetic observations are more biased relative to TRUTH. The bounded algorithms correct for this appropriately.

In sum, while the use of bounded assimilation filters does not produce significantly better or worse results in terms of the impact on modeled SIT or SND, some improvements are carried through for modeled SIC. While the full impact of boundedness in filtering is limited in this study, these filters could still provide a path to eliminating postprocessing if further infrastructure designed to simultaneously constrain SIC and categorized area in CICE-SCM-DART were developed.

3.2 Category Assimilation

More so than constraining the data assimilation with bounded filters, assimilating the model's categorized ice thickness distribution directly improves the results. First, assimilating categorized area or volume (or both) tends to lead to higher MAE reductions in modeled SIT and SIC, particularly in the cases ~~that used unbounded regression in the multivariate adjustments~~ use either fully bounded or fully unbounded filters (Fig. 6a,c). Additionally, while modeled SND is found to be degraded in nearly all cases presented here, categorized observations ~~significantly reduce the degree to which assimilating ice observations increases the MAE~~ assimilated with a fully bounded filter are found to increase pMAE by 20% (Fig. 6e).

There also appears to be evidence that assimilating categorized observations may consistently constrain the sea ice state ~~regardless of the~~ across various mean state grid cell ~~thickness~~thicknesses. In Fig. 8, assimilating SIT observations and categorized area observations perform comparably to constrain a categorized sea ice state that is relatively thick and thus has a non-negligible amount of ice in each category, including the thickest. By comparison, Fig. 10 (bottom row) presents a ~~comparison~~ case in which ~~we have restricted CICE-SCM's ability to mechanically ridge ice~~the dynamics forcing is withheld from the model integration, thereby preventing the buildup of ice in the thickest two ice categories ~~.~~via mechanical processes (i.e. ridging). In all other respects, the model configuration is identical to previous experiments. This leads to an overall thinner mean state ~~, qualitatively similar to a first-year ice regime. In this thinner ice case, SIT observations cease~~ in which SIT observations fail to constrain the thick ice categories~~(4 and 5)~~. While the erroneous adjustments made in the thickest two ice categories during assimilation are relatively minimal compared to the total grid-cell mean SIT ~~, we do find~~ (note the y-axes in Fig. 10), we observe that they lead to noticeable low biases in modeled SIC (not shown). Assimilating categorized area observations appears to avoid this issue entirely (Fig. 10, ~~top~~ bottom row)—the modeled quantities produced by doing so are consistent with TRUTH in all categories and total SIC.

There exist at least two potential applications of this ~~finding for~~ result in more realistic experiments. First, in more practical applications, the assimilation of categorized variables may avoid introducing small errors in low-concentration ITD categories that occur when assimilating SIT~~and thus~~. This has the potential to mitigate the overall error propagation of the model during intervals in which real-world SIT observations are historically unavailable to constrain the state (i.e. during summer months). Second, it has been noted in previous work that assimilating SIT can lead to biases in the sea ice edge (Riedel & Anderson, 2023), which introduces an incentive to assimilate SIC as well as SIT, despite the negative impact SIC can have on modeled quantities away from the ice edge. The consistency resulting from assimilating categorized observations in multiple ice states, including ~~thin, first-year type regimes~~ regimes in which the ice state is skewed to one end of the ITD, suggests a better solution for constraining the sea ice state everywhere in the Arctic.
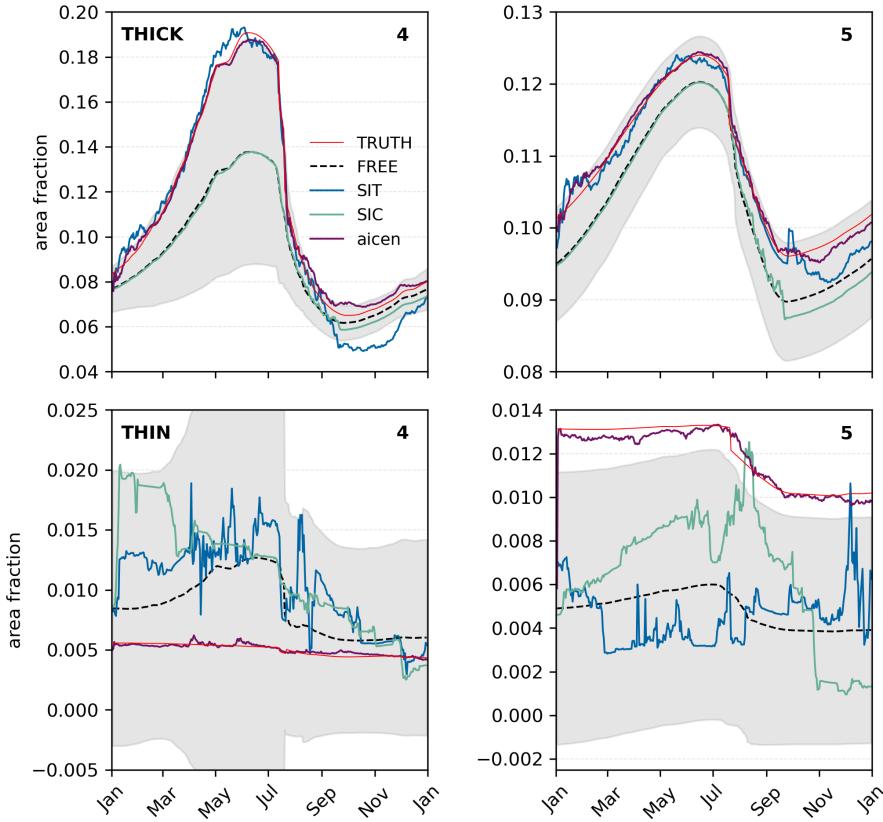
Figure 10. Category-level impact of assimilating in a thin vs thick ice state. The model's two thickest area categories for the standard case (top row, THICK) are repeated from Fig. 8. A corresponding experiment in which mechanical ridging is restricted (bottom row, THIN), leading to very low concentrations of ice in the thickest ITD categories is also shown. These results demonstrate that the strength of the relationship between SIT observations and modeled $A_{\mathrm{ice},n}$ breaks down when some categories have very little ice and that this can bias the modeled SIC result.

4 Discussion

340 This ~~study confirms~~ work reinforces the results of previous studies that assimilating SIT observations generally improves sea ice analyses over assimilating SIC observations ~~at the grid cell level~~alone. In these experiments, assimilating SIT followed by SIC leads to comparable but slightly degraded modeled quantities when compared to just assimilating SIC, which implies that for this ensemble and mean state, there is very little benefit to assimilating SIC observations, especially outside the boreal summer season. This finding, which applies to SIC analyses as well as SIT, may be due in part to the fact that we have generated

345 spread in our ensemble using only variable atmospheric forcing and that the ensemble is under-dispersive with respect to SIC for much of the year. It is worth noting, however, that in this under-dispersive SIC scenario, assimilating SIT still improves modeled SIC.
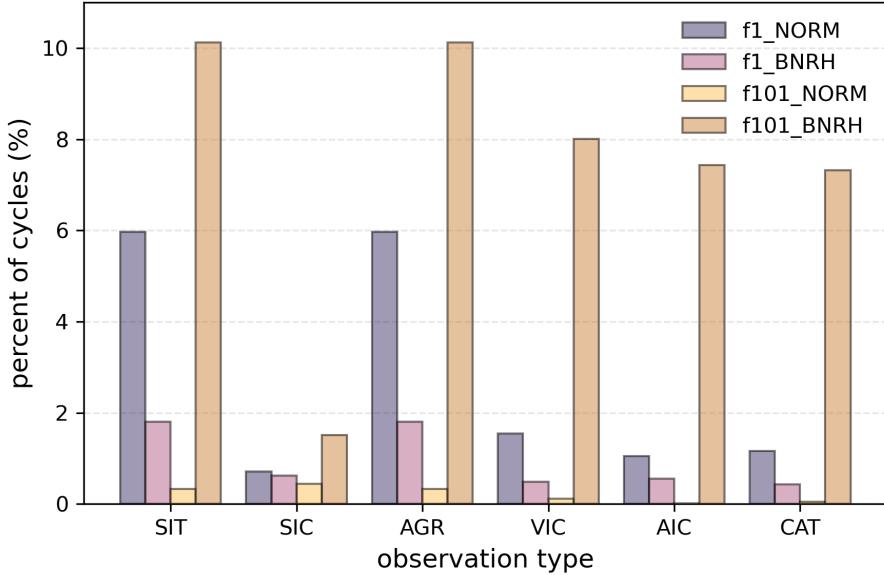
Figure 11. Percent of DA cycles requiring SIC postprocessing. For each type of observation assimilated with the four filters (f1_NORM, f1_BNRH, f101_NORM, or f101_BNRH), the percent of total DA cycles over the course of each experiment that resulted in non-physical values of aggregate SIC are shown. Note that the percent of postprocessing when assimilating SIC is artifically small, as those observations do not lead to substantial adjustments from the FREE mean except in the summertime, when grid cell aggregate SIC is decidedly lower than the upper bound (no postprocessing occurs during this time).

An emergent finding of this work is the positive impact of assimilating the categorized state (the ITD). Assimilating categorized area and volume estimates reduces MAE and increases CE on par with assimilating SIT observations, improving the model's estimates of SIC and SIT at the category level even when ~~the mean ice state is thin and~~ some categories contain very little ice. Assimilating categorized observation also reduces the forecast error beyond that of assimilating aggregate observations (Figs. 4, 8), though this is likely related at least to the fact that the categorized observation errors can be quite small (see *Methods*).

The application of a series of bounded filtering algorithms is novel to sea ice data assimilation and has highlighted the complexities of assimilating observations into a categorized distribution model such as CICE-SCM. The sometimes negligible or even detrimental impact of bounded algorithms on modeled sea ice quantities indicates a need to further tailor the CICE-SCM-DART interface such that the filters constrain categorized variables and SIC simultaneously. Bounded algorithms eliminate any need for postprocessing of $V_{\mathrm{ice},n}$, SIT, or $V_{\mathrm{sno},n}$ (not shown). However, for modeled SIC we find that bounded algorithms result in a small fraction of the adjustments made requiring SIC postprocessing (Fig. 11). Note that assimilating categorized observations reduces postprocessing requirements compared to assimilating aggregate observations, likely because the categorized observations are closer in nature to the model's state variables.

While the broad strokes of these results are expected to carry over to assimilating real-world observations, the details are likely to vary under replication in larger models, where dynamic exchange between grid cells imbues additional information into the observation-state relationships and introduces the need for localization in the data assimilation framework. We also acknowledge that the bounded filtering algorithms employed in this work depend on piece-wise distributions that are a function of the model ensemble and are relatively uninformed otherwise. DART provides the opportunity to use alternative distributions that may qualitatively shift the results. Finally, the work presented here avoids the role of various forms of model error that are present in operational data assimilation, where the observations and evolution between them are unlikely to be correctly captured by forward operators and model physics. Therefore, at the very least, the magnitude of error reductions in sea ice analyses presented may overestimate what will be achievable in more practical applications.

5   Conclusions

We have interrogated in detail the grid-cell level response of a complex sea ice model to the assimilation of various kinds of sea ice observations, including SIT, SIC, and categorized area and volume, and found that SIT and categorized observations most accurately constrain the ensemble mean forecast in both category ITD state variables and diagnostic grid-cell mean SIT and SIC; categorized observations are the only observations that perform consistently well across two different grid cell mean thickness states. Two key issues in the application of bounded data assimilation algorithms to the sea ice problem are identified. First, an approach to appropriately constrain categorized area and total SIC simultaneously is still needed. Secondly, a true understanding of where and *why* assimilation improves (or degrades) model estimates of the sea ice state depends on how well the model ensemble captures natural covariance relationships between observables and state variables ~~, a targeted study of~~ on a grid-cell scale. Quantification of these relationships requires a targeted study which is absent from previous literature ~~due to a lack of necessary observations.~~. Though observational records are short, we believe significant progress could be made in understanding the local covariance relationships between SIC and SIT with current *in situ* and remote sensing products. Future work will attempt to address the first issue and diagnose the second. Assuming that the ensemble is reasonably realistic in terms of relationship between variables, the findings presented here are expected to be qualitatively consistent in larger grid models and more practical assimilation experiments.

Code and data availability. All code used in the study can be found on Github. The CICE5 single column model is available from the CICE Consortium at https://github.com/CICE-Consortium/CICE. The Data Assmilation Research Testbed is maintained by DAReS and hosted at https://github.com/NCAR/DART. The version of DART used for this study was forked to https://github.com/mollymwieringa/DART. The python scripts and Jupyter notebooks used to configure, run, and evaluate the experiments in this study have been collected in a separate Github repository (https://github.com/mollymwieringa/cice-scm-da); the postprocessed experiment data used to produce the figures is available upon request.

Author contributions. All authors contributed to the conceptualization of the study. MMW performed the experiments, analyzed the results, and wrote the manuscript. MMW, CMB, and CPR each contributed to developing the CICE-SCM-DART interface. JLA led the development of the bounded data assimilation algorithms and consulted on their application in this study.

395 *Competing interests.* The authors declare no competing interests.

Acknowledgements. This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1755088. CMB and MMW acknowledge support from NASA ROSES Grant Number 80NSSC21K0745. MMW was also supported in part by the University of Washington College of the Environment's Integral Environmental Big Data Research Fund. The authors would like to specially thank the DAReS team at NSF NCAR/CISL,
400 Ian Grooms, Alek Petty, Jon Poterjoy, and David Bailey for many insightful conversations and helpful technical support. We thank the editor and anonymous reviewers for constructive comments that helped improve the manuscript.

References

Allard, R. A., Farrell, S. L., Hebert, D. A., Johnston, W. F., Li, L., Kurtz, N. T., Phelps, M. W., Posey, P. G., Tilling, R., Wallcraft, A. J.: Utilizing CryoSat-2 sea ice thickness to initialize a coupled ice-ocean modeling system, Adv. Space Res., 62, 1265–1280, https://doi.org/10.1016/J.ASR.2017.12.030, 2018.

Anderson, J. L.: An ensemble adjustment kalman filter for data assimilation, Mon. Wea. Rev., 129, 2884–2903, https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2, 2001.

Anderson, J. L.: A non-Gaussian ensemble filter update for data assimilation, Mon. Wea. Rev., 138, 4186-4198, 2010.

Anderson, J. L.: A marginal adjustment rank histogram filter for non-Gaussian ensemble data assimilation, Mon. Wea. Rev., 148, 3361–3378, 2020.

Anderson, J. L.: A quantile-conserving ensemble filter framework. Part I: Updating an observed variable, Mon. Wea. Rev., 150, 1061–1074, https://doi.org/10.1175/MWR-D-21-0229.1, 2022.

Anderson, J. L.: A quantile-conserving ensemble filter framework. Part II: Regression of observation increments in a probit and probability integral transformed space, Mon. Wea. Rev., *in press,* https://doi.org/10.1175/MWR-D-23-0065.1, 14 July 2023.

Anderson, J. L., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., and Arellano A.: The Data Assimilation Research Testbed: A community facility, Bull. Am. Meteorol. Soc., 90, 1283-1296, https://doi.org/10.1175/2009BAMS2618.1, 2009.

Balan-Sarojini, B., Tietsche, S., Mayer, M., Balmaseda, M., Zuo, H., De Rosnay, P., Stockdale, T., and Vitart, F.: Year-round impact of winter sea ice thickness observations on seasonal forecasts, The Cryosphere, 15, 325–44, https://doi.org/10.5194/TC-15-325-2021, 2021.

Bitz, C. M. and Roe, G. H.: A mechanism for the high rate of sea ice thinning in the Arctic Ocean, J. Clim., 17, 3623–3632, 2004.

Blockley, E. W. and Peterson, K. A.: Improving Met Office seasonal predictions of Arctic sea ice using assimilation of CryoSat-2 thickness, The Cryosphere, 12, 3419–38. https://doi.org/10.5194/TC-12-3419-2018, 2018.

Brennan, M. K., and Hakim, G.J.: Reconstructing Arctic Sea Ice over the Common Era Using Data Assimilation, Journal of Climate, 35, 1231–1247, https://doi.org/10.1175/JCLI-D-21-0099.1, 2022.

Chen, Z., Liu, J., Song, M., Yang, Q., and Xu, S.: Impacts of assimilating satellite sea ice concentration and thickness on Arctic sea ice prediction in the NCEP Climate Forecast System, J. Clim., 30, 8429–46, https://doi.org/10.1175/JCLI-D-17-0093.1, 2017.

Chevallier, M. and Salas y Melia, D.: The role of sea ice thickness distribution in the Arctic sea ice potential predictability: A diagnostic approach with a coupled GCM, J. Clim., 25, 3025-3038, https://doi.org/10.1175/JCLI-D-11- 00209.1, 2012.

El Gharamti, M., Raeder, K., Anderson, J. L., and Wang, X.: Comparing adaptive prior and posterior inflation for ensemble filters using an atmospheric general circulation model, Mon. Wea. Rev., 147, 2535–2553, https://doi.org/10.1175/MWR-D-18-0389.1, 2019.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean Dyn., 53, 343-367, 2003.

Fiedler, E.K., Martin, M.J., Blockley, E., Mignac, D., Fournier, N., Ridout, A., Shepherd, A., Tilling, R.: Assimilation of sea ice thickness derived from CryoSat-2 along-track freeboard measurements into the Met Office's Forecast Ocean Assimilation Model (FOAM), The Cryosphere, 16, 61-85, https://doi.org/10.5194/tc-16- 61-2022, 2022.

Hersbach H., Bell B., Berrisford P., et al. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146: 1999–2049. https://doi.org/10.1002/qj.3803, 2020.

Houtekamer, P. L. and Zhang, F.: Review of the ensemble Kalman filter for atmospheric data assimilation, Mon. Wea. Rev., 144, 4489-4532, 2016.

CICE Consortium Icepack: Icepack version 1.1.0, Zenodo [code], https://doi.org/10.5281/zenodo.1213462, 2020. a, b

Kalman, R. E.: A new approach to linear filtering and prediction problems, Trans. ASME, 82, 35-45, 1960.

440   Kimmritz, M., Counillon, F., Bitz, C.M., Massonnet, F., Bethke, I., and Gao Y.: Optimising assimilation of sea ice concentration in an Earth system model with a multicategory sea ice model, Tellus A: Dynamic Meteorology and Oceanography, 70, 1-23, https://doi.org/10.1080/16000870.2018.1435945, 2018.

Korosov, A., Rampal, P., Ying, Y., Ólason, E., and Williams, T.: Towards improving short-term sea ice predictability using deformation observations, The Cryosphere, 17, 4223–4240, https://doi.org/10.5194/tc-17-4223-2023, 2023.

445   Lindsay, R. W., Ice deformation near SHEBA, Journal of Geophysical Research, 107(C10), 8042, https://doi.org/10.1029/2000JC000445, 2002.

Lipscomb, W. H.: Remapping the thickness distribution in sea ice models, J. Geophys. Res. Oceans, 106, 13 989-14 000, 2001.

Lisæter, K. A., Evensen, G., and Laxon, S. W.: Assimilating synthetic CryoSat sea ice thickness in a coupled ice-ocean model, Journal of Geophysical Research: Oceans, 112, 7023, https://doi.org/10.1029/2006JC003786, 2007.

450   Massonnet, F., Fichefet, T., and Goosse, H.: Prospects for improved seasonal Arctic sea ice predictions from multivariate data assimilation, Ocean Modelling, 88, 16–25, https://doi.org/10.1016/J.OCEMOD.2014.12.013, 2015.

Mu, L., Yang, Q., Losch, M., Losa, S. N., Ricker, R., Nerger, L., and Liang, X.: Improving sea ice thickness estimates by assimilating CryoSat-2 and SMOS sea ice thickness data simultaneously, Quarterly Journal of the Royal Meteorological Society, 144, 529–38, https://doi.org/10.1002/QJ.3225, 2018.

455   Mu, L., Losch, M., Yang, Q., Ricker, R., Losa, S. N., and Nerger, L.: Arctic-wide sea ice thickness estimates from combining satellite remote sensing data and a dynamic ice-ocean model with data assimilation during the CryoSat-2 period, Journal of Geophysical Research: Oceans, 123, 7763–80, https://doi.org/10.1029/2018JC014316, 2018.

Nash, J.E. and Sutcliffe, J.V.: River flow forecasting through conceptual models, Part I: a discussion of principles, J. Hydrol., 10, 282-290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

460   Raeder, K., Hoar, T.J., El Gharamti, M., Johnson, B. K., Collins, N., Anderson, J. L., Steward, J., and Coady, M.: A new CAM6 + DART reanalysis with surface forcing from CAM6 to other CESM models, Sci Rep, 11, 16384, https://doi.org/10.1038/s41598-021-92927-0, 2021.

Riedel, C. and Anderson, J.: Exploring non-Gaussian sea ice characteristics via Observing System Simulation Experiments, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-96, 21 Mar 2023.

465   Riedel, C., Wieringa, M., and Anderson, J.: Exploring Bounded Non-parametric Ensemble Filter Impacts on Sea Ice Data Assimilation, Mon. Wea. Review [in press], (2023).

Sakov, P., Counillon, F., Bertino, L., Lister, K. A., Oke, P. R., and Korablev, A.: TOPAZ4: An ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Science, 8, 633–56, https://doi.org/10.5194/OS-8-633-2012, 2012.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R.: Uncertainty in modeled Arctic sea ice volume, Journal of
470   Geophysical Research, 116, C00D06, https://doi.org/10.1029/2011JC007084, 2011.

Thorndike, A., Rothrock, D. A., Maykut, G. A., and Colony, R.: The thickness distribution of sea ice, Journal of Geophysical Research, 80, 4501-4513, https://doi.org/10.1029/JC080i033p04501, 1975.

Williams, N., Byrne, N., Feltham, D., Van Leeuwen, P. J., Bannister, R., Schroeder, D., Ridout, A., and Nerger, L.: The effects of assimilating a sub-grid-scale sea ice thickness distribution in a new Arctic sea ice data assimilation system, The Cryosphere, 17, 2509–2532,
475   https://doi.org/10.5194/tc-17-2509-2023, 2023.

Xie, J., Counillon, F., and Bertino, L.: Impact of assimilating a merged sea-ice thickness from CryoSat-2 and SMOS in the Arctic Reanalysis, The Cryosphere, 12, 3671–91, https://doi.org/10.5194/TC-12-3671-2018, 2018.

Yang, Q., Losa, S. N., Losch, M., Tian-Kunze, X., Nerger, X., Liu, J., Kaleschke, L., and Zhang, Z.: Assimilating SMOS sea ice thickness into a coupled ice-ocean model using a local SEIK filter, Journal of Geophysical Research: Oceans, 119, 6680–92, https://doi.org/10.1002/2014JC009963, 2014.

Zhang, Y. F., Bitz, C. M., Anderson, J. L., Collins, N. S., Hendricks, J., Hoar, T. J., Raeder, K. D., and Massonnet, F.: Insights on sea ice data assimilation from perfect model Observing System Simulation Experiments, J. Clim., 31, 5911–26, https://doi.org/10.1175/JCLI-D-17-0904.1, 2018.

Zhang, Y. F., Bitz, C. M., Anderson, J. L., Collins, N. S., Hoar, T. J., Raeder, K. D., and Blanchard-Wrigglesworth, E.: Estimating parameters in a sea ice model using an ensemble Kalman filter, The Cryosphere, 15, 1277–1284, https://doi.org/10.5194/tc-15-1277-2021, 2021.

Table 1. QCEFF filter components. Includes the distributions used to represent the model ensemble in observation-space incrementing and state-space regression; the observation error distribution associated with each filter; any DART filter equivalents; and relevant references.

Filter Name	Obs. Space Dist.	State Space Dist.	Obs. Error Dist.	DART Filter Equivalent	References
f1_NORM	normal	normal	bounded normal	EAKF	Anderson (2001)
f1_BNRH	normal	BNRH	bounded normal	*none*	*none*
f101_NORM	BNRH	normal	bounded normal	BRHF	Riedel et al. (2023)
f101_BNRH	BNRH	BNRH	bounded normal	*none*	Anderson (2023)

Table 2. Assimilation experiments. Filter refers to the filter type used (see Table 1); Obs. Kind to the type of observation assimilated~~; and Obs~~. ~~Error to the formula used to determine an individual~~ Observation error ~~estimate for~~ estimates associated with each observation ~~at each timestep. a_n and v_n refer to the area and volume in the nth ITD category; $\overline{h_n}$ refers to the midpoint thickness in the same category. h_R and h_L~~ kind are ~~the upper- and lower-most thickness bounds used to define the nth ITD category.~~ shown in Table 3.

Case Name	*Filter*	*Obs. Kind* ~~*Obs. Error*~~
SIT_f1_NORM	f1_NORM	SIT ~~$\sigma_{SIT} = 0.1\,SIT$~~
SIC_f1_NORM	f1_NORM	SIC ~~$\sigma_{SIC} = -0.5(SIC^2 - SIC)$~~
AGR_f1_NORM	f1_NORM	SIT, SIC ~~$\sigma_{SIT}, \sigma_{SIC}$~~
AIC_f1_NORM	f1_NORM	aicen ~~$\sigma_{a_n} = \left(\frac{a_n}{\overline{h_n}}\right)^2 \frac{(h_R - h_L)^2}{12}$~~
VIC_f1_NORM	f1_NORM	vicen ~~$\sigma_{v_n} = (a_n)^2 \frac{(h_R - h_L)^2}{12}$~~
CAT_f1_NORM	f1_NORM	aicen, vicen ~~$\sigma_{a_n}, \sigma_{v_n}$~~
SIT_f1_BNRH	f1_BNRH	SIT ~~$\sigma_{SIT} = 0.1\,SIT$~~
SIC_f1_BRNH	f1_BNRH	SIC ~~$\sigma_{SIC} = -0.5(SIC^2 - SIC)$~~
AGR_f1_BNRH	f1_BNRH	SIT, SIC ~~$\sigma_{SIT}, \sigma_{SIC}$~~
AIC_f1_BNRH	f1_BNRH	aicen ~~$\sigma_{a_n} = \left(\frac{a_n}{\overline{h_n}}\right)^2 \frac{(h_R - h_L)^2}{12}$~~
VIC_f1_BNRH	f1_BNRH	vicen ~~$\sigma_{v_n} = (a_n)^2 \frac{(h_R - h_L)^2}{12}$~~
CAT_f1_BNRH	f1_BNRH	aicen, vicen ~~$\sigma_{a_n}, \sigma_{v_n}$~~
SIT_f101_NORM	f101_NORM	SIT ~~$\sigma_{SIT} = 0.1\,SIT$~~
SIC_f101_NORM	f101_NORM	SIC ~~$\sigma_{SIC} = -0.5(SIC^2 - SIC)$~~
AGR_f101_NORM	f101_NORM	SIT, SIC ~~$\sigma_{SIT}, \sigma_{SIC}$~~
AIC_f101_NORM	f101_NORM	aicen ~~$\sigma_{a_n} = \left(\frac{a_n}{\overline{h_n}}\right)^2 \frac{(h_R - h_L)^2}{12}$~~
VIC_f101_NORM	f101_NORM	vicen ~~$\sigma_{v_n} = (a_n)^2 \frac{(h_R - h_L)^2}{12}$~~
CAT_f101_NORM	f101_NORM	aicen, vicen ~~$\sigma_{a_n}, \sigma_{v_n}$~~
SIT_f101_BNRH	f101_BNRH	SIT ~~$\sigma_{SIT} = 0.1\,SIT$~~
SIC_f101_BNRH	f101_BNRH	SIC ~~$\sigma_{SIC} = -0.5(SIC^2 - SIC)$~~
AGR_f101_BNRH	f101_BNRH	SIT, SIC ~~$\sigma_{SIT}, \sigma_{SIC}$~~
AIC_f101_BRNH	f101_BNRH	aicen ~~$\sigma_{a_n} = \left(\frac{a_n}{\overline{h_n}}\right)^2 \frac{(h_R - h_L)^2}{12}$~~
VIC_f101_BNRH	f101_BNRH	vicen ~~$\sigma_{v_n} = (a_n)^2 \frac{(h_R - h_L)^2}{12}$~~
CAT_f101_BNRH	f101_BNRH	aicen, vicen

Table 3. Observation error estimates as a function of observation kind. Obs. Kind refers to the type of observation assimilated. Obs. Error refers to the formula used to determine an individual error estimate for each observation at each timestep. a_n and v_n refer to the area and volume in the nth ITD category; $\overline{h_n}$ refers to the midpoint thickness in the same category. h_R and h_L are the upper- and lower-most thickness bounds used to define the nth ITD category.

Obs. Kind	Obs. Error
SIT	$\sigma_{SIT} = 0.1 SIT$
SIC	$\sigma_{SIC} = -0.5(SIC^2 - SIC)$
SIT, SIC	$\sigma_{SIT}, \sigma_{SIC}$
aicen	$\sigma_{a_n} = \left(\frac{a_n}{\overline{h_n}}\right)^2 \frac{(h_R - h_L)^2}{12}$
vicen	$\sigma_{v_n} = (a_n)^2 \frac{(h_R - h_L)^2}{12}$
aicen, vicen	$\sigma_{a_n}, \sigma_{v_n}$

Appendix A
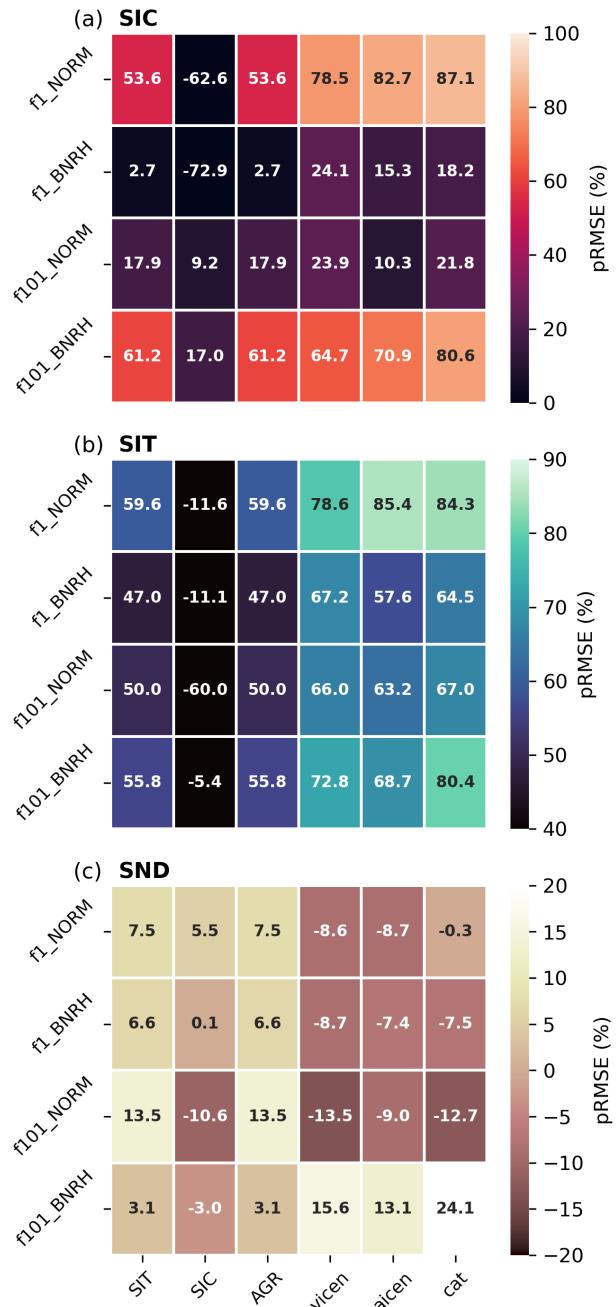
Figure A1. RMSE reduction as a function of filter type and observation kind. Same as Fig. 6 but for percent RMSE reduction (pRMSE).