



RESEARCH LETTER

10.1002/2015GL065860

Key Points:

- Forecasts from Sea Ice Outlook dynamical models are skillful at 1–2 month lead times
- SIO forecast skill is less than expected from sea ice hindcasts
- Initial-condition perturbation experiment shows importance of model physics to forecast uncertainty

Supporting Information:

- Table S1 and Figures S1–S3

Correspondence to:

E. Blanchard-Wrigglesworth,
ed@atmos.washington.edu

Citation:

Blanchard-Wrigglesworth, E., R. I. Cullather, W. Wang, J. Zhang, and C. M. Bitz (2015), Model forecast skill and sensitivity to initial conditions in the seasonal Sea Ice Outlook, *Geophys. Res. Lett.*, 42, doi:10.1002/2015GL065860.

Received 20 AUG 2015

Accepted 15 SEP 2015

Accepted article online 25 SEP 2015

Model forecast skill and sensitivity to initial conditions in the seasonal Sea Ice Outlook

E. Blanchard-Wrigglesworth¹, R. I. Cullather², W. Wang³, J. Zhang⁴, and C. M. Bitz¹

¹Department of Atmospheric Sciences, University of Washington, Seattle, Washington, USA, ²Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, USA, ³NOAA/NWS/NCEP/Climate Prediction Center, College Park, Maryland, USA, ⁴Polar Science Center, Applied Physics Laboratory, University of Washington, Seattle, Washington, USA

Abstract We explore the skill of predictions of September Arctic sea ice extent from dynamical models participating in the Sea Ice Outlook (SIO). Forecasts submitted in August, at roughly 2 month lead times, are skillful. However, skill is lower in forecasts submitted to SIO, which began in 2008, than in hindcasts (retrospective forecasts) of the last few decades. The multimodel mean SIO predictions offer slightly higher skill than the single-model SIO predictions, but neither beats a damped persistence forecast at longer than 2 month lead times. The models are largely unsuccessful at predicting each other, indicating a large difference in model physics and/or initial conditions. Motivated by this, we perform an initial condition sensitivity experiment with four SIO models, applying a fixed -1 m perturbation to the initial sea ice thickness. The significant range of the response among the models suggests that different model physics make a significant contribution to forecast uncertainty.

1. Introduction

The rapid loss of Arctic sea ice, especially during the summer, that has taken place in recent decades has resulted in growing interest in the predictability of sea ice on seasonal timescales, in part spurred by expanding socioeconomic activities in the region. Since 2008, the Study of Environmental Arctic Change has led an effort to collect and synthesize September Sea Ice Outlooks (SIOs) from the Arctic research community. The outlooks are predictions of September sea ice extent produced in late spring and early summer using a range of methods (heuristic, statistical, and dynamical models) and published on three separate dates in early June, early July, and early August. Results from SIO are collected and summarized by the Sea Ice Prediction Network (SIPN) and published on the Internet by the Arctic Research Consortium of the United States (see <http://www.arcus.org/sipn/sea-ice-outlook>).

Forecast skill in SIO over 2008–2013 was evaluated by *Stroeve et al.* [2014]. To measure skill, individual submissions to SIO were considered as single deterministic forecasts. Overall, skill was only marginally better than a linear trend forecast, and statistical predictions were found to have slightly higher skill than dynamical model predictions. Unexpectedly, forecast skill did not significantly improve as the forecast lead time decreased.

While the application of dynamical models to real-world sea ice prediction is in its infancy, considerable progress has been made over the last decade in assessing and understanding potential predictability in dynamical models, particularly fully-coupled global climate models (GCMs), using a “perfect-model” approach [e.g., *Koenigk and Mikolajewicz*, 2009; *Holland et al.*, 2010; *Blanchard-Wrigglesworth et al.*, 2011a; *Day et al.*, 2014a]. These studies agree that significant initial-value predictability is about 1–2 years for sea ice area (or extent) and 3–4 years for sea ice volume, while forced predictability (arising from external forcing) emerges about 5 years after forecast initialization. Thus, seasonal sea ice forecasting is an initial-value problem. Additionally, several studies have recently found skill in retrospective forecasts (hindcasts) of September sea ice extent from spring or early summer initializations over the recent satellite-based observation record [e.g., *Sigmond et al.*, 2013; *Chevallier et al.*, 2013; *Wang et al.*, 2013; *Peterson et al.*, 2014; *Msadek et al.*, 2014; *Guemas et al.*, 2014].

There are two primary sources of uncertainty in seasonal predictions: poorly known initial conditions and model uncertainty. In perfect-model studies, initial conditions are known perfectly and there are no model

uncertainties, in the sense that the model is used to predict itself. The chaotic growth of prescribed infinitesimal errors in the initial conditions is the only source of forecast error, and thus, the skill found in perfect-model studies is considered to be the upper limit of predictability for a specific model.

When these models are used to forecast the real world, both sources of error need to be considered. Initial conditions are often poorly constrained, and model physics are an approximation of the real world. To date, there have been few efforts to assess how potential predictability of sea ice depends on either source of error. Considering the first source of error (imperfect initial conditions), *Day et al.* [2014b] and *Msadek et al.* [2014] find that correct initialization of sea ice thickness is key for skillful predictions, while considering the second source of error (model uncertainty), *Juricke et al.* [2014] assess how different sea ice strength formulations affect sea ice predictability.

While completely attributing the gap between potential and actual forecast skill to the source of error is a complex problem, progress can be made in understanding current forecast skill. In this paper we first assess the skill of dynamical models in SIO and consider how it compares with the skill from all published perfect-model and hindcast experiments, and that of a simple statistical benchmark. To help isolate the role of model uncertainty on forecast error, we then show results from an initial-condition perturbation experiment using four models that have submitted to SIO in the past. In this experiment, conducted as part of the 2014 SIPN international workshop, we assess the response of the models to an identical perturbation in the sea ice thickness initial conditions that were used for the 2013 SIO.

2. SIO Model Forecast Skill

We begin by assessing the forecast skill of the model contributions to SIO. The root-mean-square error (RMSE) is used to evaluate forecast skill of September sea ice extent for outlooks from 2009 to 2014 in each submission period (early June, early July, and early August, Figure 1). The RMSE is shown for both the ensemble mean of each model and the multimodel ensemble mean (considering the ensemble mean of each model as one member of the multimodel ensemble), which we refer to as $RMSE_{sio-obs}$ and $RMSE_{mean(sio)-obs}$ respectively. They are defined as

$$RMSE_{sio-obs} = \sqrt{\frac{1}{N} \sum_{i=2009}^{2014} \sum_{j=1}^J (y_{ij} - x_i)^2} \quad (1)$$

and

$$RMSE_{mean(sio)-obs} = \sqrt{\frac{1}{N} \sum_{i=2009}^{2014} (\bar{y}_i - x_i)^2}, \quad (2)$$

where y_{ij} are mean forecasts of individual models, x_i is the observed September sea ice extent, \bar{y}_i is the multimodel ensemble mean of forecasts, j is the model, i is the year, J is the total number of forecasts for each submission call, and N is the total number of variables in the summation. We use the National Snow and Ice Data Center September sea ice extent values [Fetterer et al., 2002, updated 2014] for x_i in equations (1) and (2). For all years, SIO model forecasts total 35, 43, and 37 for the June, July, and August submission dates, respectively.

SIO forecasts are generally initialized in the weeks preceding the submission date, and this time-varying initialization is illustrated by the line plots in Figure 1. To compare with the results of SIO models, we add to Figure 1 the RMSE of previously published perfect-model studies, the RMSE of bias-corrected and detrended hindcast experiments, and the RMSE for all SIO submissions (including statistical and heuristic methods from *Stroeve et al.* [2014]). We also include the RMSE of a damped anomaly persistence forecast, whereby the anomaly from the linear trend in the month preceding the forecast is applied to the following September linear trend value, scaled by the autocorrelation coefficient between both months and the ratio of the standard deviations of both months [Van den Dool, 2006]. We also show $\sqrt{2}\sigma$, where σ is the standard deviation of observed detrended September sea ice extent—this value represents the lower threshold of skill [Collins, 2002] and, by definition, is also the predictability offered by a linear trend forecast, while a RMSE value of zero represents perfect skill. We apply an F test to calculate the degree of statistical significance of the difference among various RMSE values. Details of all experiments displayed in Figure 1 are shown in Table 1.

While other metrics are commonly used in predictability studies, such as the anomaly correlation coefficient, or ACC (see Collins [2002] for a discussion of different metrics), we choose to use RMSE since it arguably

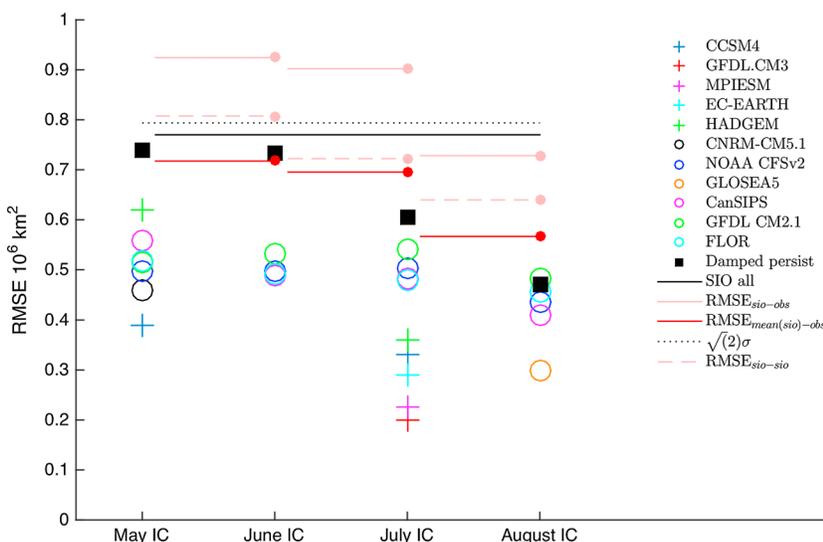


Figure 1. RMSE of September sea ice extent forecasts. Crosses show values from perfect-model predictability experiments; circles indicate results from hindcasts using dynamical models (see Table 1 for details). “SIO all” is the RMSE value for all SIO forecasts reported in *Stroeve et al. [2014]*, $RMSE_{sio-obs}$ is the RMSE of all individual SIO model forecasts, $RMSE_{mean(sio)-obs}$ is the RMSE of the SIO model-mean forecast, and $RMSE_{sio-sio}$ is the RMSE of the SIO ensemble considered as a perfect model, where one model integration is taken as the truth and all other models’ integrations are forecasts. The time-varying initialization of SIO models in the weeks preceding SIO calls is illustrated by plotting lines instead of points.

provides a more physically intuitive estimate of predictability error—an actual sea ice extent value rather than a correlation coefficient. Moreover, calculating an ACC requires a minimum sample size of $n = 3$ for each model in SIO and would thus eliminate from our sample size those models that have only one or two yearly submissions, which represent 35% of the total sample size.

Perfect-model forecasts tend to have higher skill than the hindcasts, yet the skill in both beats the skill threshold from damped persistence. Interestingly, the skill of the hindcasts improves little at shorter lead times—much less than the improvement in skill of damped persistence. Conversely, individual SIO model forecasts ($RMSE_{sio-obs}$) exhibit very high RMSE values and hence do not beat damped persistence or even $\sqrt{2}\sigma$ for the June and July initializations, indicating no skill. The $RMSE_{sio-obs}$ is significantly different from all hindcast RMSE values at the 99% level. SIO model skill is slightly higher when considering the SIO model-mean predictions ($RMSE_{mean(sio)-obs}$) yet only marginally beats the damped persistence threshold for June initialization. However, given the insignificant correlation of sea ice extent anomalies between May and

Table 1. List of Dynamical Models Shown in Figure 1

Model	Experiment Type	Period	Reference
GFDL CM3	perfect model	present-day control run	<i>Tietsche et al. [2014]</i>
HadGEM1.2	perfect model	present-day control run	<i>Tietsche et al. [2014]</i>
MPI-ESM-LR	perfect model	present-day control run	<i>Tietsche et al. [2014]</i>
EC-EARTH V2	perfect model	present-day control run	<i>Tietsche et al. [2014]</i>
NCAR CCSM4	perfect model	present-day control run	<i>Blanchard-Wrigglesworth et al. [2011a]</i>
CanSIPS	hindcast	1979–2009	<i>Sigmond et al. [2013]</i>
CNRM-CM5.1	hindcast	1990–2009	<i>Chevallier et al. [2013]</i>
MetOffice GLOSEA5	hindcast	1996–2009	<i>Peterson et al. [2014]</i>
NOAA CFSv2	hindcast	1982–2007	<i>Wang et al. [2013]</i>
GFDL CM2.1	hindcast	1982–2013	<i>Msadek et al. [2014]</i>
GFDL CM2.5 FLOR	hindcast	1982–2013	<i>Msadek et al. [2014]</i>
SIO models	forecast	2009–2014	<i>Stroeve et al. [2014]</i>

September [Blanchard-Wrigglesworth *et al.*, 2011b], damped persistence is not expected to yield significant skill at this lead time. The $\text{RMSE}_{\text{mean}(\text{sio})-\text{obs}}$ is not significantly different from the hindcast RMSE values at the 95% level.

We can also explore the SIO model ensemble by evaluating it as a perfect model—in other words, by considering one of the model integrations as the truth and all other models' integrations as forecasts. We call this $\text{RMSE}_{\text{sio-sio}}$ in Figure 1 and define it as

$$\text{RMSE}_{\text{sio-sio}} = \sqrt{\frac{1}{N} \sum_{i=2009}^{2014} \sum_{j=1}^J \sum_{k \neq j} (y_{ki} - y_{ji})^2}, \quad (3)$$

where the notation is as that in equation (1). This calculation of RMSE of the SIO model ensemble is equivalent to the way perfect-model RMSE is calculated in earlier works [see Blanchard-Wrigglesworth *et al.*, 2011a].

This metric can shed light on whether there is a consistent bias (in model physics and/or initial conditions) among models, as a small value would indicate small intermodel differences in initial conditions and physics, whether the models skillfully simulate observations or not. If the models predict each other better than observations, this would be evidence of common errors in the physics and initial conditions employed across the models. In the case that a common bias exists across models, one would expect $\text{RMSE}_{\text{sio-sio}}$ to be lower than $\text{RMSE}_{\text{sio-obs}}$. However, if these RMSE values are equal, then it would indicate the models have the same skill (or lack of skill) in predicting each other or observations. Figure 1 shows that $\text{RMSE}_{\text{sio-sio}}$ is only marginally lower than $\text{RMSE}_{\text{sio-obs}}$, and not statistically significantly different at the 95% level, indicating that SIO models are as poor at predicting each other as they are poor at predicting observations. This result suggests that there are significant differences in physics and/or initial conditions across SIO models.

3. Sensitivity of SIO Models to Initial Conditions

Forecast systems may produce forecasts that disagree with one another by using different initial conditions and/or different model physics. Dynamical models that participate in SIO range from regional ice-ocean models to global fully-coupled models, while the observations that are assimilated and the ice-ocean reanalyses that are used as initial conditions vary substantially (M. Chevallier, personal communication, 2015). In order to investigate the role of model uncertainty on forecast error and spread, we next test the sensitivity of the forecast to sea ice thickness initial conditions of dynamical models in SIO. To achieve this, prior to the 2014 SIO workshop modeling groups that submitted to SIO in 2013 were invited to perform an initial-condition perturbation experiment in which the sea ice thickness was reduced by 1 m relative to the initial conditions used for the 2013 SIO—referred to as the control run hereafter.

Four groups performed the experiment with the following models: National Center for Atmospheric Research (NCAR) Community Climate System Model version 4 (CCSM4) [Gent *et al.*, 2011], NASA Global Modeling and Assimilation Office (GMAO) Goddard Earth Observing System 5 Atmosphere-Ocean General Circulation Model (AOGCM) [Suarez *et al.*, 2008], NOAA Climate Forecast System version 2 (CFSv2) [Saha *et al.*, 2014], and Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS) [Zhang and Rothrock, 2003]. For the SIO forecast, NCAR CCSM4 utilizes a global fully-coupled GCM initialized with sea ice thickness anomalies taken from PIOMAS. To create an ensemble forecast, initial atmospheric states are varied across ensemble members by taking consecutive days in the control run centered around the initialization date. NASA GMAO and NOAA CFSv2 are global fully-coupled seasonal forecasting systems. PIOMAS is a regional ice-ocean model that assimilates sea ice concentration and is forced by NOAA/National Centers for Environmental Prediction reanalysis data. In forecast mode, PIOMAS produces a seven-member ensemble, each member with prescribed atmospheric conditions taken from reanalysis of the previous 7 years. Further details on the models and the methodology employed for the experiment and SIO can be found in Table S1 in the supporting information and online at the SIO URL (<http://www.arcus.org/search-program/seaiceoutlook/2013/july>). NCAR CCSM4, NASA GMAO, and NOAA CFSv2 are initialized in early May, while PIOMAS is initialized in early June. Initial sea ice thickness was reduced in all four models by 1 m, except in regions where the original sea ice thickness is less than 1.5 m. In these regions, the perturbation was reduced linearly as a function of the original control sea ice thickness, as $p = h/1.5$, where p is the perturbation and h is the control ice thickness. Initial sea ice area and extent was kept unchanged from the control run. The choice of a 1 m sea ice thickness perturbation is arbitrarily chosen to be large enough to provide a high signal-to-noise response in the experiments and is a 2–4 σ anomaly of sea ice thickness [Blanchard-Wrigglesworth and Bitz, 2014].

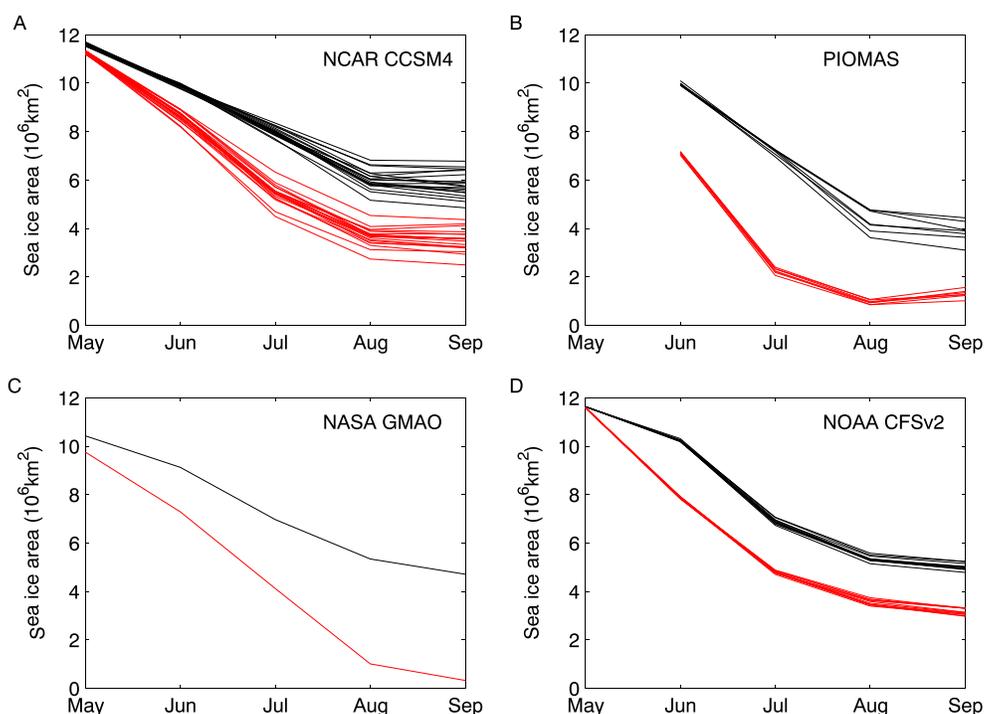


Figure 2. Monthly sea ice area in 10^6 km^2 in the control (SIO 2013, black) and experiment (red) runs for (a) NCAR CCSM4, (b) University of Washington (UW) PIOMAS, (c) NASA GMAO, and (d) NOAA CFSv2.

Figure 2 shows monthly sea ice area in all models for the control and perturbation forecasts. In all four models, sea ice area is smaller relative to the control, as would be expected from the reduced sea ice thickness in the initial conditions. However, the response varies across all four models throughout the forecast integration period (May through September for NCAR CCSM4, NASA GMAO, and NOAA CFSv2, and June through September for PIOMAS). Figure 3a shows the mean monthly sea ice area difference between the perturbed and control forecasts. In NCAR CCSM4 the perturbed forecast loses $2.5 \times 10^6 \text{ km}^2$ relative to the control between May and July, and subsequently, the loss in sea ice area stabilizes. NASA GMAO follows a similar trajectory until July, but subsequently, the loss in sea ice area in the perturbed forecast relative to the control increases to almost $4.5 \times 10^6 \text{ km}^2$ by August—indeed, NASA GMAO loses most sea ice by August (see Figure 2c). NOAA CFSv2 has a rapid initial loss of $2\text{--}2.5 \times 10^6 \text{ km}^2$ by June, which then stabilizes for the remainder of the forecast, while PIOMAS has a strong initial loss in sea ice area over the first 2 months (June–July), as its summer sea ice area becomes strongly reduced (see Figure 2b). By September, the range of response across all four models is a loss of $1.9 \times 10^6 \text{ km}^2$ to $4.4 \times 10^6 \text{ km}^2$. It is unclear why the models have such a wide range of response, and as shown in Figure S3, the range is only weakly related to the range of sea ice thickness initial conditions.

We also explore how the growth of the forecast ensemble spread of individual models (a measure of potential predictability) responds to the initial condition perturbation (see Figure 3b). In NCAR CCSM4, the spread of the perturbed forecast ensemble for June and July grows faster than the control forecast ensemble, indicating a more rapid loss of potential predictability in the perturbed forecasts in these months. However, in August and September both ensemble spreads are comparable (not significantly different at the 95% level), indicating that potential predictability is similar. PIOMAS has a different response—while the control forecast ensemble spread grows comparably to that of the control forecast ensemble in NCAR CCSM4, the perturbed forecast ensemble has a much more reduced spread after July, which is visually apparent in Figure 2b. Lastly, NOAA CFSv2 has a much slower growth in its control forecast ensemble spread than either NCAR CCSM4 or PIOMAS, while its perturbed forecast ensemble spread is not significantly different from the control forecast ensemble spread. Thus, all three models that performed a multiple-run perturbation forecast exhibit different responses in the growth of forecast ensemble spread.

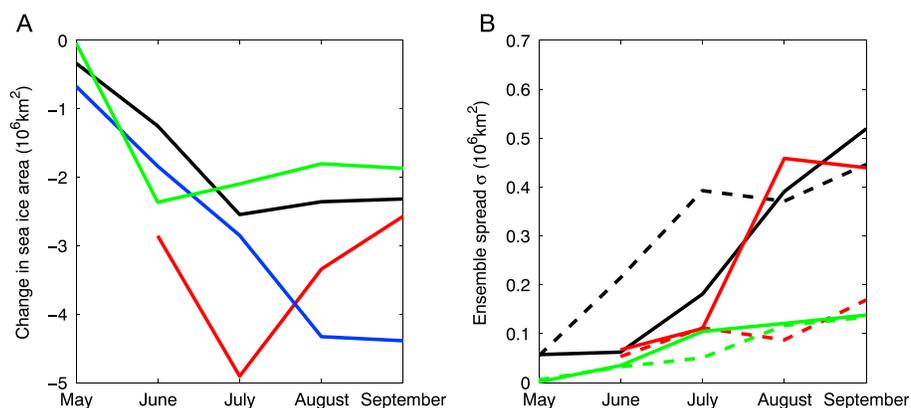


Figure 3. (a) Monthly sea ice area difference in 10^6 km² between the perturbed and control forecast ensemble-mean sea ice area for NCAR CCSM4 (black), UW PIOMAS (red), NASA GMAO (blue), and NOAA CFSv2 (green) and (b) forecast ensemble spread standard deviation in the control (solid) and experiment (dashed) simulations for NCAR CCSM4 (black), UW PIOMAS (red), and NOAA CFSv2 (green).

4. Discussion

The low skill in SIO models relative to the hindcasts is surprising, particularly considering that some of the models that have produced hindcasts have also submitted outlooks to SIO (e.g., NOAA CFSv2 and MetOffice GLOSEA5). It is possible that summer sea ice extent has been less predictable in recent years compared to previous decades (the time period for which most of the hindcasts were produced, see Table 1), a result also found within the Geophysical Fluid Dynamics Laboratory (GFDL) hindcasts [Msadek *et al.*, 2014]—indeed, thinner sea ice conditions may lead to lower predictability [e.g., Holland *et al.*, 2010] and may offset the expected forecast improvement offered by improved data availability for initialization in recent years [Msadek *et al.*, 2014]. One may also consider how sea ice persistence has evolved throughout the satellite era, since sea ice predictability is intrinsically linked with persistence of sea ice anomalies [Day *et al.*, 2014a], and it has recently been shown that the persistence and reemergence of sea ice anomalies show significant interannual variability [Bushuk *et al.*, 2015]. Thus, reduced persistence throughout the summer may be indicative of reduced potential predictability. However, neither persistence nor the skill offered by damped persistence has decreased in recent years (Figure S1).

Interestingly, some climate reanalysis systems that are used to provide initial conditions for hindcast and forecast experiments show reduced fidelity in simulating observed September sea ice extent variability in recent years compared to the earlier period of the satellite observation record (Figure S2) [Msadek *et al.*, 2014, Figure 1b], and it is plausible that this would imply a reduction in forecast or hindcast skill during recent years. It is worth noting that the RMSE of SIO submissions of both NOAA CFSv2 and MetOffice GLOSEA5 models is significantly higher than their hindcast RMSE values shown in Figure 1, and no better than $RMSE_{sio-obs}$ (not shown). Care must be taken, however, when interpreting this result given the small sample size of individual model submissions to SIO (10 and 7 submissions during 2011–2014 from NOAA CFSv2 and MetOffice GLOSEA5, respectively).

The multimodel mean of SIO forecasts offer slightly higher skill than the single-model ensemble prediction, a typical feature seen in other areas of seasonal climate prediction [e.g., Hagedorn *et al.*, 2005] yet neither beats significantly a damped persistence forecast, perhaps helping explain why statistical models show higher skill than dynamical models in SIO [Stroeve *et al.*, 2014]. Furthermore, models are equally unskilled at predicting each other, indicative of a large difference in model physics and/or initial conditions.

Four dynamical models that have participated in SIO have significantly different responses to identical initial condition perturbations. This result implies that different model physics can result in significantly different forecast sensitivity to identical initial condition perturbations. Thus, model physics are likely responsible for a significant component of the large spread in the multimodel September SIOs. Improving the skill of September sea ice predictions will thus depend not only on better observations and assimilation of the models' initial conditions but also on further improving model physics.

Acknowledgments

We thank Michael Sigmond, Bill Merryfield, Woo-Sung Lee, and Rym Msadek for making the CanSIPS, NOAA CFSv2, GFDL CMOR, and ECDA hindcast skill scores available and Larry Hamilton for assistance with the SIO data sets. We thank two anonymous reviewers for helpful suggestions on how to improve the paper. The SIO data in the paper are available at <http://www.arcus.org/sipn/sea-ice-outlook>. EBW and CMB were supported by ONR grant N0014-13-1-0793. JZ was supported by ONR grant N00014-12-1-0112.

The Editor thanks two anonymous reviewers for their assistance in evaluating this paper.

References

- Blanchard-Wrigglesworth, E., and C. M. Bitz (2014), Characteristics of Arctic sea-ice thickness variability in GCMs, *J. Clim.*, *27*(21), 8244–8258.
- Blanchard-Wrigglesworth, E., C. M. Bitz, and M. H. Holland (2011a), Influence of initial conditions and climate forcing on predicting Arctic sea ice, *Geophys. Res. Lett.*, *38*, L18503, doi:10.1029/2011GL048807.
- Blanchard-Wrigglesworth, E., K. Armour, C. M. Bitz, and E. deWeaver (2011b), Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations, *J. Clim.*, *24*(1), 231–250, doi:10.1175/2010JCLI3775.1.
- Bushuk, M., D. Giannakis, and A. J. Majda (2015), Arctic sea ice reemergence: The role of large-scale oceanic and atmospheric variability, *J. Clim.*, *28*, 5477–5509.
- Chevallier, M., D. Salas y Méliá, A. Voltaire, M. Déqué, and G. Garric (2013), Seasonal forecasts of the pan-Arctic sea ice extent using a GCM-based seasonal prediction system, *J. Clim.*, *26*(16), 6092–6104.
- Collins, M. (2002), Climate predictability on interannual to decadal time scales: The initial value problem, *Clim. Dyn.*, *19*, 671–692.
- Day, J., S. Tietsche, and E. Hawkins (2014a), Pan-Arctic and regional sea ice predictability: Initialisation month dependence, *J. Clim.*, *27*(12), 4371–4390.
- Day, J., E. Hawkins, and S. Tietsche (2014b), Will Arctic sea ice thickness initialization improve seasonal forecast skill?, *Geophys. Res. Lett.*, *41*, 7566–7575, doi:10.1002/2014GL061694.
- Fetterer, F., K. Knowles, W. Meier, and M. Savoie (2002, updated 2014), Sea ice index. [Available at <http://nsidc.org/data/go2135.html>.]
- Gent, P., et al. (2011), The community climate system model version 4, *J. Clim.*, *24*(19), 4973–4991.
- Guemas, V., et al. (2014), A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales, *Q. J. R. Meteorol. Soc.*, doi:10.1002/qj.2401.
- Hagedorn, R., F. J. Doblas-Reyes, and T. Palmer (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept, *Tellus A*, *57*(3), 219–233.
- Holland, M. M., D. A. Bailey, and S. Vavrus (2010), Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3, *Clim. Dyn.*, *36*, 1239–1253, doi:10.1007/s00382-010-0792-4.
- Juricke, S., H. F. Goessling, and T. Jung (2014), Potential sea ice predictability and the role of stochastic sea ice strength perturbations, *Geophys. Res. Lett.*, *41*, 8396–8403, doi:10.1002/2014GL062081.
- Koenig, T., and U. Mikolajewicz (2009), Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model, *Clim. Dyn.*, *32*, 783–798, doi:10.1007/s00382-008-0419-1.
- Msadek, R., G. Vecchi, M. Winton, and R. Gudgel (2014), Importance of initial conditions in seasonal predictions of Arctic sea ice extent, *Geophys. Res. Lett.*, *41*, 5208–5215, doi:10.1002/2014GL060799.
- Peterson, K. A., A. Arribas, H. Hewitt, A. Keen, D. Lea, and A. McLaren (2014), Assessing the forecast skill of Arctic sea ice extent in the GloSea4 seasonal prediction system, *Clim. Dyn.*, *44*, 147–162.
- Saha, S., et al. (2014), The NCEP climate forecast system version 2, *J. Clim.*, *27*(6), 2185–2208.
- Sigmond, M., J. Fyfe, G. Flato, V. Kharin, and W. Merryfield (2013), Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system, *Geophys. Res. Lett.*, *40*, 529–534, doi:10.1002/grl.50129.
- Stroeve, J. C., L. Hamilton, C. M. Bitz, and E. Blanchard-Wrigglesworth (2014), Predicting September sea ice: Ensemble skill of the SEARCH sea ice outlook 2008–2013, *Geophys. Res. Lett.*, *41*, 2411–2418, doi:10.1002/2014GL059388.
- Suarez, M. J., et al. (2008), The GEOS-5 data assimilation system—documentation of versions 5.0.1, 5.1.0, and 5.2.0, *Tech. Rep. Ser. on Global Model. and Data Assimilation*, NASA/TM-2008-10460627, NASA/Goddard Space Flight Center, Greenbelt, Md.
- Tietsche, S., J. Day, V. Guemas, W. Hurlin, S. Keeley, D. Matei, R. Msadek, M. Collins, and E. Hawkins (2014), Seasonal to interannual Arctic sea-ice predictability in current GCMs, *Geophys. Res. Lett.*, *41*, 1035–1043, doi:10.1002/2013GL058755.
- Van den Dool, H. (2006), *Empirical Methods in Short-Term Climate Prediction*, Oxford Univ. Press, Oxford, U. K.
- Wang, W., M. Chen, and A. Kumar (2013), Seasonal prediction of Arctic sea ice extent from a coupled dynamical forecast system, *Mon. Weather Rev.*, *141*(4), 1375–1394.
- Zhang, J., and D. Rothrock (2003), Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates, *Mon. Weather Rev.*, *131*(5), 845–861.