J Day et al. Response to reviewers comments on "The Arctic Predictability and Prediction on Seasonal-to-Interannual TimEscales (APPOSITE) data set"

We would like to thank the reviewers for taking the time to carefully read this paper and for some very useful suggestions. Whilst we agree that this dataset is ideal for some of the additional analyses suggested by the reviewers and that these would be very informative. As the APPOSITE project has come to an end, we would like to point out that the primary role of this manuscript is to provide a descriptive reference for this dataset, so that it is well described for future use. Therefore our primary action in response to the comments has been to clarify and expand on the description of the experiment and archived data, where suggested by each reviewer. That said, we have taken the time to follow a suggestion by both reviewers to examine the initial state dependence of sea ice predictability and have included a new subsection and additional figure on this point.

Reviewer 1:

In this paper a multi-model protocol for analysing potential model predictability is introduced, focusing on the potential predictability of the Arctic sea ice conditions on the seasonal to interannual timescale. The setup of the ensemble simulations is explained as well as the diagnostics used to analyse potential predictability of Arctic sea ice extent and volume. Seven different models have contributed to create a dataset following the basic guidelines of this protocol, with some difference in the more specific details such as ensemble size and number of ensemble start dates. The results for the ensembles of four of these models regarding potential Arctic sea ice predictability have previously been discussed in a paper by Tietsche et al. (2014), while the results for the remaining three models are added to the discussion for this paper.

In general I appreciate the effort of the authors to make the data available to the broader scientific community and to use this publication as a reference for the setup of the experiment protocol. Analysing potential predictability and the differences therein between GCMs is certainly an important area of research, especially as a tool to inform seasonal prediction systems of the feasibility of future improvements. The paper is generally well written and the structure is straight forward. While I appreciate the authors' choice to keep this publication short and concise, I do have some comments that might increase the length of the paper quite a bit. My main point of critique is that the paper is very close to the previous publication by Tietsche et al. (2014) without presenting a more detailed description of the experimental setup, and without discussing the new results equally detailed as the previous study. Since both aspects are the main points of this paper, they should be extended, still keeping them as separate aspects of the same publication, i.e. first the discussion of the protocol, then the application to the newly contributed models, highlighting the importance of both.

General comments

As a first comment and to repeat my question of the summary, could the authors be more specific regarding the focus of this paper and how it differs from the Tietsche et al. (2014) publication. I assume you want to equally focus on the results for the additional three models as well as on the general setup of the protocol. But at the moment I would claim that both parts are a bit too short and not very detailed.

In the Abstract and Introduction, we have been more specific about the goals of the analysis, which is to provide an updated estimate the predictability forecast horizon for sea-ice extent and volume also mentioning the additional work on sea ice extent and volume predictability initial state dependence.

Some more specific examples regarding the experimental setup:
When you write about the high, low and medium sea ice states used for initialisation, how is that reflected in the actual ensemble start dates? Does this relate to the sea ice volume, the sea ice area or average sea ice thickness? Are they separated in some way in the archiving structure? Are you trying to estimate the impacts of different initial conditions by this approach, even though some models only have 8 different start dates, which would make it difficult to actually assess differences in the predictability caused by the initial state?
Choosing the start dates was essentially left up to the participating group, but we encouraged them to sample a range of initial states based on pan-arctic extent and volume. The aim was to investigate state dependence of sea ice predictability. These points are made explicit in Sec 2.2.

When you say "well spaced" (page 8815, line 18) how is this defined? Was there a minimum spacing between successive start dates that you have generally defined for all models to insure independence of the initial state?
As the modelling centres chose their own start dates there is a bit of a range, the minimum spacing is 3 years for GFDL, but longer for other models.

How was the length of the control run defined? Different models have different spin up times and might take longer to equilibrate. After only 100 years I wouldn't think any model has really equilibrated, as can be seen by the strong drift of most of the models.
Could you comment on some of these details, stating advantages and disadvantages of the choices you had to make to generate this dataset. Also, in this context, the time axis for the panels in Figure 1 doesn't make much sense to me. The start date of each model control seems more or less random, even though the text reads they started from (the same?) static state oceanic depth profile.
Again, the particulars of initialisation, spinup and length of control were dependant on the modelling groups. Some groups had a 1990/Present day simulation with their model, but many did not. As this is not part of the CMIP5 DECK so groups either had to create the necessary boundary conditions and start a fresh simulation. As no groups outside Reading were funded to do these simulations it was difficult to standardise this approach. However, since every model has been spunup for at least 100 years, intermodal differences in climatology are unlikely to be affected significantly by these differences. Since we are looking at initial value predictability only over the first three years, it is unlikely that issues such as drift play a large role in the assessment of predictability. We have expanded the text in this section to make this more apparent.

It's worth noting that even after 1000s of years, many climate models still drift, so this is something we have to live with. In practice, even with 200 years of model time series, models with pronounced low-frequency variability can exhibit apparent drifts even when they are in "equilibrium" purely due to the particular phases of variability captured in the window used for trend analysis.

Effectively the times in Figure 1 are random since this is just the model clock year in the control run. We only show the period of the model that was used to calculate the reference climate mean and *standard deviation. The spinup period of the models was not collected from the centres or archived. We have made these points explicit in the Figure 1 caption and Section 2.1 text.*

Were the SST perturbations applied globally, also in areas of sea ice cover?
Yes, we make it explicit that they were applied at all ocean cells.

Regarding the two metrics, were they applied to detrended monthly means? If so, was the detrending based on the control or all ensemble members? It would simplify the explanations for the metrics if you would actually expand the expectation value as was done in Collins (2002), also to show which normalization you chose (what is sigma?).
This is the standard deviation of the model climate, as shown in Figure 4. We have made this clearer in the text.

What kind of significance test was applied to the ACC?
We used a T-test, details are now given in the text.

Are there any specific plans to extend this dataset, i.e. to include more models? Or to use this dataset for other predictability studies?
There are plans for this, but they are dependent on the outcome of funding proposals. We think these plans are too tentative to be worth mentioning in the text.

Some more specific examples regarding the results:

The sea ice models in this study differ in many aspects. Could you comment a bit on how this affects the results? For example, do models with similar albedo and melt pond parametrizations produce similar results, or do models with similar sea ice dynamics (number of sea ice classes and so on) produce similar mean states and climate variability? I know this is a difficult questions, since the other model components show significant differences as well. However, it would be interesting to know whether some systematic differences can be identified.
As the reviewer states, this is a difficult question to answer. All we can say is that we have not identified any such links between sea ice model formulation and other properties. We believe a more targeted experiment would be required to say more about this.

Could you please expand the paragraph about the mean state and climate variability. For one, it is not surprising that the mean states of the models are different compared to the mean state of the observations, which have been recorded over a shorter period of time and under transient forcing conditions.

We have added to the discussion here. However, because this is simply designed to highlight the variety in model climate states rather than robustly assess the realism of each model, we do not present a detailed assessment of model climate. This aim is also made explicit in the text.

Furthermore, could you comment on how model variability and mean state affect the predictability metrics.

This is an important question, however with the number of models available we only have 7 data points to derive any relationships, which we believe is too few to do anything robustly. We are planning to extend this dataset as part of a later proposal and come back to this point. We also include this as an open question in the conclusions section.

What are the consequences of the different drifts in the models? Do you expect a more equilibrated model to provide a more accurate estimate of potential predictability?

We have taken account of this in the metrics used by using a time varying climatology in the case of ACC. This is explained in more detail in the text.

Why didn't you apply any of the spatial predictability metrics which were used by Tietsche et al. (2014)? What about the other start dates provided, especially January? Since the extended results of this paper are mentioned as one of the two major contributions of this study, it would be nice if the paragraphs about the model results (page 8818) could be expanded, providing more details on the differences and similarities in predictability between the models and possible reasons for that.

We have added a paragraph to the end of Section 3.2 to discuss some of the open questions relating the predictability to climate and some potential next steps. We have also clarified that we are extending the analysis of Tietsche et al. in particular to assess the limit of extent and volume predictability from July. Hence we do not utilise the Jan predictions, or the spatial measures.

Page 8818, lines 12-15: How does this relate to the results of the current study?

Have added 'Indicating that the winter sea ice extent predictability horizon may be significantly beyond the 3 years simulated in these experiments' to the end of this sentence.

Page 8818, line 23: There is always a chance that you remove internal variability by detrending, also for a longer timeseries. It is just less likely.

Have added "is likely to significantly", the point being that it will be enough to significantly affect the predictability metric."

Page 8818, lines 26-27, and page 8819, lines 1-3: This paragraph is difficult to read. Maybe you could break up the sentences.

This paragraph has been rewritten.

Page 8819, lines 6-7: The differences of the mean state and variability between models and observations wasn't discussed in any detail.

I have changed this to say we have presented the mean state and variability.

Page 8819, line 17: Not really true for E6F (early loss of predictability for sea ice volume; no re-emergence of predictability for NRMSE).

This statement is less true for E6F, we have changed this "Sea ice volume is ***generally*** *more predictable than sea ice extent"*

Minor comments:

Page 8811, line 16: Change to “Unprecedented”, “opportunities”, “businesses”.

Page 8811, line 17: Change to “but has also”.

Page 8811, line 23: “appreciation”.

All above changed

Page 8812, line 1: What do you mean by “significantly skillful”? Could you also give a reference here?

Changed to “have statistically significant skill”

Page 8812, lines 9-11: Please rephrase this sentence. Be more specific about this “fundamental limit”, which has different timescales for the atmosphere and the sea ice.

Done

Page 8812, lines 20-21: Please expand this. What are the disadvantaged of potential predictability studies? How does model uncertainty affect predictability estimates?

We have added some additional discussion here.

Page 8813, line 5: Change to “: : : climate variables as well. In order: : :”.

Changed as suggested

Page 8813, line 10: Differences in design such as?

Page 8813, line 12: Differences in the results such as?

Have rewritten this section on motivations.

Page 8813, lines 13-16: Again, could you name some of the differences, either here or before?

OK

Page 8814, line 22: Change to “sea ice”.

Done

Page 8815, line 1: Change to “distribution, as well as”.

Done

Page 8815, lines 11-13: Can you quantify this/be more specific? Does this have consequences for summer sea ice predictability when it comes to different model mean states?

Added as stated above

Page 8815, line 20: Change to “depending on”.

Done

Page 8816, line 8: Remove comma at the end.

Done

Page 8816, line 21: Change to “inter-model”.

Done

Page 8818: Mention Figure 5 again, after first sentence of 3.2 and 3.3.
Page 8819, line 14: Change to “interannual”.
Page 8820, line 7: Change to “constraints:”.
Page 8820, lines 8-11: Could you give a reference here?
Page 8820, line 23: Change to “submodel&frequency”.
Page 8820, line 23 onwards: Check for text size and font here and on the next page.
Page 8820, line 25: Is it “1” (this line) or “r1” (next page, line 1).
Figure 2 and 3: Is the average taken over the entire simulation length or only for the years after the spin-up?
Figure 4: Mention detrending in caption.

All Done as suggested.

Anonymous Referee #2

The manuscript presents an updated version of the APPOSITE dataset that is originally presented and discussed in Tietsche et al 2014 and Day et al 2014. In its current version, the manuscript adds unfortunately little new information or insights into sea ice predictability to these two papers, and I feel it is, as it stands, a missed chance to use the dataset to explore issues that are at present topical in the field. I would encourage the authors to extend their analysis.

Since publication of Tietsche et al. (2014), the APPOSITE protocol was followed by a number of additional models and this database has been made openly available as a community resource. This is why we believe that it is useful to publish an extend the description of the dataset and update the results of Tietsche et al. We agree that there are still many open questions in this area, which is why we have made the effort to make this data openly available. It provides a unique resource to investigate initial value predictability in multiple models.

I suggest below a few ideas to explore. How does predictability depend on mean state? The APPOSITE dataset, with its start dates split between high,medium, and low initial conditions (p8815 L18), is currently the best opportunity to explore this question. If you find that the number of ensembles/runs is still not large enough to yield statistically robust results, this finding would still be useful for the community - I suspect the answer will depend on whether in fact there are (meaningful) inherent differences in predictability with mean state. Given current trends in sea ice in observations, exploring this issue is key.

How can we understand the inter-model differences in predictability? While the patterns in change of predictability with time are similar across models (e.g., predictability barrier in SIV in early summer, winter>summer SIE predictability in years 2,3), there is a considerable spread in predictability across models as you point out in the conclusions (as an aside, I would guess given your ensemble size that the inter-model differences are significant, but it would be good to calculate and show this). This is a significant result. I note that in Day et al 2014 (Jclim), you explore links between predictability and persistence, and persistence and mean state. It would be good to do this with the current larger dataset. Are models with higher predictability more ’persistent’ (Figure

1 shows models have varying degrees of persistence in their control runs)? It has been shown (B-W and Bitz, 2014) that models with thicker sea ice tend to have longer thickness persistence timescales - does this help explain inter-model differences? By looking at Figure 4 and 5, it's hard to figure out if there's a link between total volume and predictability. Perhaps a scatter plot of e.g., mean NRMSE over year 1 against mean SIV would help. (You could even split each model into its 3 high/medium/low ICs and obtain 6*3 datapoints).
We agree that the question of how predictability depends on model mean state, or other properties of model climate is a crucial one. However, we feel that given the limited set of models it will be difficult to infer any robust relationships. However as part of a follow-up proposal we intend to extend these runs to other models so that such an analysis will be possible.

We have however extended our analysis in this work to investigate how initial value predictability depends on whether the model is in a high, medium or low state at its initial state. This is in a separate section of Section 3 (3.4).

Can you extend the dynamic v thermodynamic analysis of Tietsche et al 2014 (see their section 3.3, Fig3) to more models? Discerning which physical process leads to loss of predictability, particularly at seasonal timescales, would be an important result. Additionally, considering if the relative importance of different processes varies between different initialization seasons (January vs July) would be equally insightful.

Unfortunately the diagnostics required to perform this analysis were not available for models other than MPI and HadGEM.

Minor:

There are several spelling mistakes - please proof read cautiously

We have thoroughly proofread the document and removed a number of spelling mistakes.

Manuscript prepared for Geosci. Model Dev. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LATEX class copernicus.cls.
Date: 22 January 2016

The Arctic Predictability and Prediction on Seasonal-to-Interannual TimEscales (APPOSITE) data set

J. J. Day[1], S. Tietsche[2], M. Collins[3], H. F. Goessling[4], V. Guemas[5,6], A. Guillory[7], W. J. Hurlin[8], M. Ishii[9], S. P. E. Keeley[2], D. Matei[10], R. Msadek[6], M. Sigmond[11], H. Tatebe[12], and E. Hawkins[1]

[1]NCAS-Climate, Department of Meteorology, University of Reading, Reading, UK
[2]European Centre for Medium-Range Weather Forecasts, Reading, UK
[3]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
[4]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[5]Institut Català de Ciències del Clima, Barcelona, Spain
[6]CNRM/GAME, Toulouse, France
[7]British Atmospheric Data Centre, Rutherford Appleton Laboratory, Chilton, UK
[8]Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA
[9]Meteorological Research Institute, Tsukuba, Japan
[10]Max Planck Institute for Meteorology, Hamburg, Germany
[11]Canadian Centre for Climate Modelling and Analysis, Environment Canada, Victoria, Canada
[12]Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

Correspondence to: J. J. Day (j.j.day@reading.ac.uk)

Abstract

Recent decades have seen significant developments in ~~seasonal-to-interannual timescale~~ climate prediction capabilities at seasonal-to-interannual timescales. However, until recently the potential of such systems to predict Arctic climate had ~~not~~ rarely been assessed. This paper describes a multi-model predictability experiment which was run as part of the Arctic Predictability and Prediction On Seasonal to Inter-annual Timescales (APPOSITE) project. The main goal of APPOSITE was to quantify the timescales on which Arctic climate is predictable. In order to achieve this, a coordinated set of idealised initial-value predictability experiments, with seven general circulation models, was conducted. This was the first model intercomparison project designed to quantify the predictability of Arctic climate on seasonal to inter-annual timescales. Here we present a description of the archived data set (which is available at the British Atmospheric Data Centre)~~and an update of the project's results.~~ , an assessment of Arctic sea ice extent and volume predictability estimates in these models, and an investigation into to what extent predictability is dependent on the initial state.

The inclusion of additional models expands the range of sea ice volume and extent predictability estimates, demonstrating that there is model diversity in the potential to make seasonal-to-interannual timescale predictions. We also suggest that sea ice forecasts started from extreme high and low sea ice initial states exhibit higher levels of potential predictability than forecasts started from close to the models mean state.

Although designed to address Arctic predictability, we describe the archived data here so that others can use this data set ~~could also be used~~ to assess the predictability of other regions and modes of climate variability on these timescales, such as the El Niño Southern Oscillation.

1 Introduction

~~Unprecednteted~~ Unprecedented climate change in the Arctic has opened up ~~oportunities for buniness~~ opportunities for business in diverse sectors such as fossil fuel and mineral extraction, shipping and tourism but has also put pressure on local communities, who are dependent on the ice for their livelihoods (Emmerson and Lahn, 2012; Stephenson et al., 2013). The need for these stakeholder groups to avoid hazardous sea ice and weather conditions has increased demand for Arctic sea ice forecasts at seasonal-to-interannual time scales ~~(Eicken, 2013)~~(Eicken, 2013; Jung et al., 2016). These local interests and a growing ~~apreciation~~ appreciation of the importance of the Arctic in mid-latitude weather phenomena (Jung et al., 2014) have motivated the development of seasonal sea ice prediction systems (e.g. Sigmond et al., 2013; Chevallier et al., 2013; Wang et al., 2013; Peterson et al., 2014) which are initialised from observations.

It has previously been shown that these sea ice prediction systems ~~are significantly skillful at~~ exhibit significant skill in predicting summer sea ice ~~cover~~extent a season ahead (Guemas et al., 2014), but diagnosing the source of forecast errors is problematic~~(Guemas et al., 2014)~~. Forecast errors may be due to both inadequate representation of important physical processes in the model ~~(e.g. melt ponds, Schröder et al., 2014)~~ (such as melt ponds, Schröder et al., 2014) or inadequate knowledge of initial-state ~~vector variables~~conditions, such as sea ice thickness (Day et al., 2014a; Msadek et al., 2014; Massonnet et al., 2015), which is not currently used to initialise operational forecasts. Sea ice predictability is also inherently limited due to chaotic, unpredictable atmospheric variability (Blanchard-Wrigglesworth et al., 2011b; Holland et al., 2010) which will lead to irreducible errors in sea ice predictions at seasonal and longer timescales, fundamentally limiting the timescale at which sea ice will be predictable (Tietsche et al., 2016). If the skill of a given forecast system is already close to this fundamental limit it will not be possible to further increase the leadtime at which the forecast is ~~skillful~~skilful.

To determine if there is the potential to improve the operational prediction systems, we consider a more ~~idealized~~ idealised situation. The "perfect-model" approach to estimating predictability involves producing initial-value ensemble-predictions with a General Circulation Model (GCM), which are verified against the model itself rather than against observations of the real world (following Griffies and Bryan, 1997b). It is therefore not hampered by changes to the observational network over time or changes in predictability due to secular climate change, which hampers this kind of analysis in the real world ~~. It therefore provides an upper bound for~~ (Collins, 2002). Such studies provide an estimate of the predictive skill obtainable in a world ~~governed by the same physical equations as the model (Hawkins et al., 2015), though may not necessarily be~~ with a perfect model and complete observations. However, such estimates are not necessarily an upper bound for the limit of predictability in the real world ~~(Eade et al., 2014; Shi et al., 2015).~~ because important predictability mechanisms may be missing (Eade et al., 2014). There is an ongoing discussion in the literature on this point (e.g. Shi et al., 2015).

The perfect model approach has previously been used to quantify and understand predictability of coupled modes of climate variability, such as the Atlantic Meridional-Overturning Circulation (AMOC) (e.g. Griffies and Bryan, 1997a; Collins, 2002; Pohlmann et al., 2004) and the El Niño Southern Oscillation (ENSO) (Collins et al., 2002), leading to the development of operational seasonal-to-decadal prediction systems based on atmosphere-ocean climate models (e.g. Smith et al., 2007; Jin et al., 2008).

Using this approach Collins et al. (2006) demonstrated that the timescale on which the AMOC is predictable varies from model to model. These inter-model differences in predictability arise because different GCMs have different representations of the underlying physical equations and parameters. It is therefore likely that there will be inter-model differences in predictability for other climate variables ~~, so in order to assess uncertainty in model based estimates of the limit of predictability~~ so it is important to conduct such analyses in multiple GCMs. The APPOSITE model intercomparison was designed to diagnose the limit of initial-value predictability of Arctic sea ice in multiple GCMs. Previous studies had estimated this limit in individual climate models, but with ~~different experiment design~~slightly different experiment designs (such as Blanchard-Wrigglesworth et al., 2011b; Holland et al., 2010; Koenigk and Mikolajewicz; All these experiments demonstrated initial-value sea ice predictability on seasonal-to-interannual timescales~~but with significant differences in the details (Blanchard-Wrigglesworth et al., 2011b; Holland et al., 2010; Koenigk and Mikolajewicz, 2009; Ti~~

~~However, because the experimental protocol was~~ , however because they focussed on slightly different variables, averaging periods and because the experimental protocols were inconsistent between the studies, it was not clear whether ~~differences in predictability were inherent in the models themselves or due to differences in the experimental set-up~~the results of these studies were consistent (Guemas et al., 2014). For the APPOSITE ensemble a consistent protocol was followed ~~so that~~ to ensure that it was possible to intercompare models, so that any differences in predictability were only the result of differences in the inherent predictability of the models themselves. The first results of this project were presented in Tietsche et al. (2014).

~~Here we present~~ The primary aim of this manuscript is to provide a detailed description of the APPOSITE experiment, archived at the British Atmospheric Data Centre (BADC) (Day et al., 2015)~~and an update on the results of~~ . We also present an updated assessment of the limit of Arctic sea ice extent and volume predictability, initially presented in Tietsche et al. (2014), including more models than available at the time of ~~publication.~~ this publication. In addition we consider an open question in Arctic prediction: to what extent is sea ice predictability state dependent? In this study we consider whether sea ice extent and volume predictability is different when initialised from high and low states compared to states close to the model climatology.

The paper is outlined as follows: Sect. 2 describes the experiment in detail as well as the mean state of the models used, Sect. 3 includes an update of the results of Tietsche et al. (2014) and the state dependence analysis, followed by the conclusions in Sect. 4. Additional details of the data set, archived at the BADC, are included as Appendix A.

2 Description of the simulations

Seven different coupled climate models performed simulations for APPOSITE (see Table 1). Six of these models followed the same experimental protocol, which is described in Sect. 2.1 and 2.2. ~~One~~ For practical reasons one model, CanCM4, followed a slightly different protocol which is described in Sect. 2.3.

2.1 Control simulations

Predictability of the climate system changes with mean climate ~~(DelSole et al., 2014; Holland et al., 2010)~~ (DelSole et al., 2014) complicating the assessment of predictability in a transient climate. ~~The~~ This is likely to be particularly acute in the Arctic where the sea ice climate changes rapidly in transient simulations (Holland et al., 2010). The APPOSITE experimental protocol therefore asked for both control simulations and ensemble predictions to be conducted in GCMs with forcing fixed at present-day values.

Since the perfect-model approach uses initial conditions generated by the model itself, present-day control simulations with each model were run under fixed present-day radiative forcings. For practical reasons the year that the forcings correspond to differ between models, either 1990, ~~but by no more than a decade or two~~ 2000 or 2005 depending on the model (see Table 1). ~~Appart~~ Apart from MPI-ESM, which was initialised from year 2005 of the CMIP5 historical simulation, all other models were initialised in a static state from present day ocean temperature and salinity profiles (e.g. Conkright et al., 2002). ~~After a spin-up period of about~~ The period of spinup varied from model to model but is at least 100 ~~years, each model is~~ years. Each model was integrated for at least 100 ~~more~~ further years to fully sample the model's ~~mean state, the remaining climate~~climate, drift, and the models internal variability. ~~It~~ Data from the spinup period of each model was not archived. However, it is worth noting that despite more than a century of spinup, some of these simulations still have significant drifts in the mean sea ice ~~climatology (see Figs~~extent and volume timeseries (see Fig. 1~~and 2).~~). These drifts are accounted for by the predictability metrics we use in Section 3 and are not expected to significantly influence the estimate of predictability.

All of the models are ~~full atmosphere-ocean-seaice~~ coupled atmosphere-ocean-sea ice GCMs and each has a fully prognostic sea ice component. These account for ~~changes~~ variations in sea ice due to both thermodynamic and advective processes that result from stress internal to the sea ice as well as through interaction with the atmosphere and ocean.

Like all components of the GCMs, the sea ice models have both structural and conceptual differences. The most significant of which are their treatment of sea ice dynamics, ~~like~~ such as the local ice thickness distribution, as well as vertical heat flux through the ice ~~,~~ and heat exchange at the ice-ocean interface. Except for HadGEM1.2, E6F and MIROC5.2 the versions of the models used were those submitted to the Coupled Model Intercomparison Project Phase 5 (CMIP5). These models have been well tested and evaluated against observations and their strengths and weaknesses are well-documented (see references in Table 1). However, in order to ~~understand~~ facilitate understanding of the differences in sea ice predictability, we ~~focus on~~ present the differences in their sea ice mean state and variability.

~~The~~ Although not designed to robustly assess the realism of each model's climate this analysis shows that sea ice mean state and variability in the control runs differ considerably from model-to-model and to the observations (see Figs. 2, 3 and 4). Before calculating the standard deviation, shown in Fig. 4, a linear trend was removed from sea ice extent and volume timeseries for each model. ~~Interannual variability~~ The wide range of sea ice climates in GCMs is well known (e.g. Arzel et al., 2006; Flato et al., 2013), however the wide model variety in inter-annual variability exhibited by the different models is likely to be just as important for the determining the inherent predictability exhibited by each model. Indeed looking across the models, the inter-annual variability of summer sea ice extent in each model appears to be negatively correlated to its mean, in line with previous studies (Goosse et al., 2009; Holland et al., 2008). This does not appear to be the case for winter. It should also be noted that whilst the climate of each model is very well sampled here (over 100 years), the observational timeseries, at a length of 35 years, is much shorter.

2.2 Ensemble predictions

To diagnose the inherent predictability in each of these models, we performed a suite of ensemble predictions. The number of start dates selected from the control run differs from model to model and ranges between 8 and ~~18. These were chosen to sample~~ 18, depending on the resource limitations of each modelling centre. Whilst participating groups were responsible for choosing their own start dates, they were encouraged to pick them so that a range of high, low and medium sea ice ~~states, while keeping~~ extent and volume states were captured, in order that any dependence of sea ice predictability on the size of the initial state anomaly could be assessed (see Section 3.4). They were also encouraged to keep start dates well spaced in time~~to consider them~~ , so that they could be considered independent (see Fig. 1). The minimum spacing between start dates is 3 years in the case of GFDL-CM3, and longer in other models.

For each start date an ensemble of between 8 and 16 members was generated, ~~dependingon the model~~again depending on the resource limitations of each modelling centre. The initial conditions were taken from the control run ~~and each~~ of each model and each ensemble member differs only by a perturbation to the sea surface temperature field. ~~This perturbation~~ The perturbation used to generate the ensemble takes the form of randomly-generated spatially-uncorrelated ~~Gaussian noise,~~ noise, applied to each grid cell. This noise is sampled from a Gaussian distribution with a standard deviation of 10^{-4} K. Each ensemble member starts with a slightly different realisation of this noise. Such a perturbation is so small that it is equivalent to assuming perfect knowledge of the initial conditions. For a given start date, differences in the evolution of each ensemble member are solely determined by the chaotic nature of the simulated climate system. ~~Each~~ Note that different initialisation methods, such as lagged atmospheric conditions may lead to slightly different predictability estimates (see Hawkins et al., 2015). For each start date the ensemble was run for 3 years, with the exception of MIROC5.2, which was run for 3.5 years.

A minimum contribution for models to be included in the APPOSITE experiment was to submit a control run and predictability experiments started on the 1st July, which allows an assessment of seasonal predictions of the late-summer sea ice conditions, when the sea ice is at ~~it's~~ its lowest extent, and human activity in the the Arctic Ocean is largest. Although we restrict our analysis to the simulations started in July, some groups have also submitted simulations started in January, May and November (see Table 1 for details). Note that operational predictions are more commonly started in May. We decided to start our simulations later due to the presence of an early summer ~~,~~ predictability barrier, which might lead to a sharply decreased skill in predicting the late-summer sea ice extent minimum (Blanchard-Wrigglesworth et al., 2011a; Day et al., 2014b).

2.3 CanCM4 transient experiments

The set of simulations with the CanCM4 model use a different protocol, in order to facilitate direct comparison of these simulations with the CanSIPS operational seasonal prediction system, which uses the same climate model (Sigmond et al., 2013).

The CanCM4 simulations were different in two key respects. Firstly, they were run under a transient climate, with observed historical forcing agents prescribed. Secondly, initial-value ensembles were generated every year and only run for 1 year. In all other regards, such as the method of ensemble generation, these simulations are the same as the other APPOSITE perfect model simulations.

3 Perfect model intercomparison

An intermodel comparison of Arctic sea ice predictability, using four climate models, was published in Tietsche et al. (2014). Here we present an update of this study, including the MIROC5.2, E6F and CanCM4 climate models.

3.1 Metrics

Two predictability metrics, as defined by Collins (2002), were used to quantify predictability in this study. These make use of the fact that in a perfect model study, such as this, any ensemble member may be chosen as “the truth” or “the forecast”. Therefore it is possible to increase the effective sample size by taking each member as “the truth” in turn, and comparing it with every other member as “the forecast”. ~~The~~ For each model the Normalised Root

Mean Squared Error (NRMSE) compares forecast RMSE to the climatological variability:

$$\text{NRMSE} = \frac{\sqrt{\langle (x_{kj} - x_{ij})^2 \rangle_{i,j,k \neq i}}}{\sqrt{2\sigma^2}} \tag{1}$$

where $\langle \cdot \rangle_i$ denotes the expectation value, to be calculated by summing over the specified index with appropriate normalization, $x_{ij}(t)$ is the sea ice extent at lead time t for the ith member of the jth ensemble. The σ in the denominator is the standard deviation of the control run for the appropriate month, calculated from the whole archived timeseries (shown in Fig. 1) after the linear trend has been removed (values shown in Fig. 4). The value of the denominator is equivalent to the climatological RMSE between two independent realisations~~. Significance of this is~~, which is the limit that the RMSE term in the nominator will approach over time. Therefore the NRMSE will approach a limit of 1. The model is said to show significant predictability when the NRMSE is significantly lower than 1, as calculated using an ~~f test~~F-test, following Collins (2002).

The second metric is the anomaly correlation coefficient (ACC). This is defined as:

$$\text{ACC} = \frac{\langle (x_{ij} - \mu_j)(x_{kj} - \mu_j) \rangle_{i,j,k \neq j}}{\langle (x_{ij} - \mu_j)^2 \rangle_{i,j}}. \tag{2}$$

where μ_j is the climatological mean at the time of the jth ensemble prediction. The anomalies are calculated relative to a time varying climatology to take into account any drifts in the control run, otherwise ACC values for models with larger drifts would be biased high. For the jth start date, the climatology μ_j is the value of the linear fit at the corresponding point in the control run timeseries at the corresponding point in time. Note that we chose to use the whole timeseires for each model (after the spinup period), shown in Fig 1, to estimate the reference climate. For a detailed discussion on the impact of such choices on the estimate of predictability see Hawkins et al. (2015).

At some lead-time, both of these metrics become insignificantly different from their asymptotic limit (0 for ACC and 1 for NRMSE), and the lead-time at which this happens can be used to define the limit of predictability. ~~However, it~~ For each lead-time, significance is calculated using an F-test or t-test in the case of the NRMSE and ACC metrics respectively, where for each model the degrees of freedom used in the test is the number of start dates multiplied by the number of ensemble members run for that model. It appears that the NRMSE metric is more conservative than the ACC metric and becomes ~~insignificant~~ insignificantly different from its limit at an earlier ~~lead time~~ lead-time (see Fig. 5). Thus using both metrics gives some spread in the estimate of the time when the limit of predictability is actually reached.

3.2 Fixed forcing experiments

Although sea ice extent predictability decreases rapidly during the first year, with the exception of EC-Earth, all models (and both metrics) show significant levels of predictability for the first year ~~.~~ (see Fig. 5). After the first year of simulation, two of the models, MIROC5.2 and GFDL-CM3, show significant levels of predictability at all later lead times. At the other end of the predictability spectrum, E6F is only intermittently predictable after the first year. Predictability in E6F (and to a lesser extent HadGEM1.2) has a strong seasonal cycle with months surrounding the winter extent maximum significantly predictable until the end of the simulation and no significant summer predictability after the first year.

Sea ice volume is much more predictable than sea ice extent in all models. Apart from E6F all models exhibit significant predictability in all 3 years of the simulations. In a prognostic predictability analysis with decadal simulations, Germe et al. (2014) similarly found that winter sea ice extent was predictable out to seven years in their model, compared to three years in summer and found that volume was predictable out to nine years ahead. It is therefore likely that the winter sea ice extent predictability horizon may be significantly beyond the 3 years simulated in these experiments.

3.3 CanCM4 transient experiments

Both the NRMSE and ACC metrics indicate lower levels of predictability in CanCM4 for sea ice extent and sea ice volume ~~.~~ (see Fig. 5). It is possible that the CanCM4 model actually has inherently lower levels of initial-value predictability than the other models. However, there are reasons to expect that both metrics will ~~be more conservative using the transient protocol~~indicate lower levels of predictability not because of inherently lower levels of initial-value predictability, but because of using the shorter control run associated with the transient protocol employed by CanCM4.

In the case of NRMSE, detrending a short timeseries ~~reduces~~ is likely to significantly reduce the climatological variance, since without multiple ensemble members to estimate the forced trend, some internal variability is removed in attempting to remove the forced trend (see Hawkins et al., 2015).

~~In the case of ACC , the~~ We believe that the ACC values are lower than the estimates of other models for the following reason. The reference climate (which is a linear fit to the control run) is a much ~~closer fit~~ better fit to the data, with lower residuals, in the case of the short CanCM4 transient control run than it is for the long fixed forcing control runs~~, which have~~. This is because, in general, the long control runs have have large decadal anomalies ~~. This will reduce the correlation and is analogous to the way that the ACC between two timeseries is reduced by removing the trend from both.~~which are not well approximated by a linear fit. Therefore the simulations will exhibit lower persistence CanCM4 than would be found if the same model had been run in the fixed forcing setup, simply as a result of differing accuracy of the linear fit in each case.

3.4 State dependence of predictability

As mentioned in Section 2.2, start dates for the ensembles were chosen to sample low, medium and high sea ice extent and volume states in each model's control run. In order to estimate whether starting in different positions of model state space has an impact on skill we calculated the anomaly correlation metric again but only selecting start dates according to if they were started from a month of the control run with a low, medium or high state. This was done for most models by choosing the two lowest states, two highest states or two states closest to the mean of the control runs. E6F had 3 start dates in each class and CanCM4 had 7 in each, as a result of these models having more start dates than other models. In general, the high states are larger than 0.8 standard deviations above the mean and the low states lower than 0.8 standard deviations below the mean. To assess the start date dependence of sea ice extent predictability the start dates were binned by sea ice extent and to assess the dependence of volume predictability they were binned by volume. The ACC was recalculated for each of these bins (see Fig. 6).

Fig 6. provides a clear indication that there is indeed some start date dependence. In the case of sea ice extent, the ACC of ensembles started from years close to climatology drops very rapidly during the first 6 months of the simulations, both in the multi-model mean and in individual models (apart from HadGEM), compared to the high and low cases where ACC values stay higher for longer. The differences are most apparent in the months immediately following September, which is when freeze-up begins following the summer minimum. It may be that there are differences at longer lead times, but with this small sample size the time series of ACC are noisy and difficult to interpret.

Sea ice volume also exhibits much less predictability when initialised from states where the volume is close to the model climatology. Indeed the multi-model mean ACC falls dramatically in the medium case compared to the low and high years. Skill remains comparatively low during the rest of the simulation.

We believe the inter-model agreement over the features we highlight provide a strong indication that initialising forecasts from extreme model states of the results in more skilful forecasts of both sea ice extent and volume. Physically, one reason for this might be is that autumn and winter heat loss acts as a strong negative (stabilising) feedback. If anomalous atmospheric forcing leads to a large negative anomaly in September ice extent or thickness one year, there will also be large oceanic heat losses during the following freeze-up season areas of open water and thin ice which encourage ice production (Serreze and Stroeve, 2015). One might expect the evolution from states where this feedback dictates large heat flux anomalies to be more predictable than others. However, this behaviour might also be expected from simple arguments based on the positive auto-correlation of sea ice on these timescales. Since the sea ice extent and volume auto-correlation is positive, one might expect large anomalies to persist, leading to increased predictability when initialising from extreme states. A more in depth study in this area would be needed to differentiate between these two hypotheses.

4 Conclusions

We have presented the experimental protocol for the APPOSITE Arctic sea ice predictability multi-model intercomparison, and described the archive of model simulations which contributed to it. The mean state and variability of Arctic sea ice cover in the models was presented and compared to observed estimates~~and~~. We utilise this database to assess the limit of initial-value Arctic sea ice extent and volume predictability ~~was estimated~~ from each of the models, updating the results of Tietsche et al. (2014) to include three more models.

The results of this analysis of perfect model predictability can be summarised as follows:

- The winter sea ice extent is predictable at interannual timescales (or possibly longer timescales) in all models.
- There is significant intermodel spread in the timescale at which summer sea ice extent is predictable, with some models not showing any ~~interanual~~ interannual or longer timescale predictability, and others showing significant predictability throughout all months of the 3 year simulations.
- Sea ice volume is ~~much~~ generally more predictable than sea ice extent.

~~The data used in this study~~

Further, because prediction ensembles were started from high, medium and low sea ice states we were able to assess the state dependence of sea ice predictability for the first time. We found that for both volume and extent, the future evolution of the climate appears to be more predictable when started from high or low states compared to those forecasts started from states close to the model mean.

These data are archived at the BADC (Day et al., 2015) ~~.~~ and have been used in a number of sea ice predictability studies. These have: (i) quantified the predictability horizon for Arctic sea ice forecasts (Tietsche et al., 2014, and this study), (ii) demonstrated the existence of a spring “predictability barrier” for sea ice predictions (Day et al., 2014b), (iii) highlighted the development of sea ice thickness initialisation as a crucial step towards skilful seasonal predictions (Day et al., 2014a), (iv) quantified the sources of irreducible forecast error in Arctic predictions (Tietsche et al., 2016), and (v) been used to investigate the initial state dependence of sea ice predictability (this study). This dataset has therefore helped fill key knowledge gaps in sea ice prediction research.

However, important questions on Arctic sea ice predictability still remain. For example, a clear understanding of why predictability varies from model to model and to what extent it depends on the models mean climate remains elusive. We feel that it will be necessary to expand this set of predictability experiments in order to answer this question robustly. We hope that by making these data available, other researchers will be able to utilise them to answer these and other open questions.

As well as enabling the results of the APPOSITE project to be reproduced ~~, this will also allow these predictability experiments to~~ and allowing the community to utilise these simulations for Arctic sea ice research, this archive could also be further utilised to improve understanding of predictability of other variables on seasonal-to-interannual timescales, such as Antarctic sea ice cover (e.g. Holland et al., 2013) or even ENSO (e.g. Collins et al., 2002).

Appendix A: Database description

APPOSITE requested a specific set of variables from participants focused on sea ice analysis, but many other variables have been archived besides. The file and directory naming convention, followed by the archived data set, is very similar to that followed by CMIP5 (http://cmip-pcmdi.llnl.gov/cmip5/output_req.html).

APPOSITE required participants to prepare their data files so that they meet the following constraints.

- Data files are in netCDF file format and ideally conform to the climate and forecast (CF) metadata convention (outlined on the website http://cf-pcmdi.llnl.gov). In instances where it was not possible to produce fully CF compliant netCDF files, participants were required to follow the CMOR variable naming convention.

- There must be only one output variable per file.

- The file names have to follow the file naming convention outlined below.

Each variable is contained in a single directory of a directory tree with the following structure:

<model>/<runtype>/<submodel&frequency>/<variable>

Where runtype is "ctrl" or "pred" for the control run or ensemble predictions respectively, model is the name of the climate model (e.g. hadgem1_2, mpiesm, ...), variable is the CMOR name for a given climate variable and submodel&frequency indicates the model sub-component and frequency (e.g. Amon, Aday, Omon and Oday).

Files are named using the following convention:

<variable>_<submode&frequency>_<model>_<runtype>_<run>_<time>.nc

Where run is a concatenated string including the start year, prediction start month and ensemble member number for ensemble predictions (e.g. 2005Jul3); or simply contains "1" for a control run.

For example,

tas_Amon_hadgem1_2_ctrl_r1_200501-200512.nc for control runs,

or

tas_Amon_hadgem1_2_pred_2005Jul3_200507-200806.nc for the 3rd ensemble member of an ensemble started on the 1 July 2005.

Acknowledgements. This work was supported by the Natural Environment Research Council (grant NE/I029447/1). Helge Goessling was supported by a fellowship of the German Research Foundation (DFG grant GO 2464/1-1). Data storage and processing capacity was kindly provided by the British Atmospheric Data Centre (BADC). Thanks to Yanjun Jiao (CCCma) for his assistance with the CanCM4 simulations and to Bill Merryfield for his comments on a draft of the paper.

References

Arzel, O. and Fichefet, T. and Goosse, H.: Sea ice evolution over the 20th and 21st centuries as simulated by current AOGCMs, Ocean Modelling, 12, 3-4, 401–415, doi:10.1016/j.ocemod.2005.08.002, 2006.

Blanchard-Wrigglesworth, E., Armour, K. C., Bitz, C. M., and DeWeaver, E.: Persistence and Inherent Predictability of Arctic Sea Ice in a GCM Ensemble and Observations, J. Climate, 24, 231–250, doi:10.1175/2010JCLI3775.1, 2011a.

Blanchard-Wrigglesworth, E., Bitz, C., and Holland, M.: Influence of initial conditions and climate forcing on predicting Arctic sea ice, Geophys. Res. Lett, 38, L18503, doi:10.1029/2011GL048807, 2011b.

Chevallier, M., Salas y Mélia, D., Voldoire, A., Déqué, M., and Garric, G.: Seasonal Forecasts of the Pan-Arctic Sea Ice Extent Using a GCM-Based Seasonal Prediction System, J. Climate, 26, 6092–6104, doi:10.1175/JCLI-D-12-00612.1, 2013.

Collins, M.: Climate predictability on interannual to decadal time scales: the initial value problem, Clim. Dynam., 19, 671–692, doi:10.1007/s00382-002-0254-8, 2002.

Collins, M., Frame, D., Sinha, B., and Wilson, C.: How far ahead could we predict El Niño?, Geophys. Res. Lett., 29, 130-1–130-4, doi:10.1029/2001GL013919, 2002.

Collins, M., Botzet, M., Carril, A. F., Drange, H., Jouzeau, A., Latif, M., Masina, S., Otteraa, O. H., Pohlmann, H., Sorteberg, A., Sutton, R., and Terray, L.: Interannual to decadal climate predictability in the North Atlantic: a multimodel-ensemble study, J. Climate, 19, 1195–1203, 2006.

Conkright, M. E., Locarnini, R. A., Garcia, H. E., O'Brien, T. D., Boyer, T. P., Stephens, C., and Antonov, J. I.: World Ocean Atlas 2001: Objective analyses, data statistics, and figures: CD-ROM documentation, US Department of Commerce, National Oceanic and Atmospheric Administration, National Oceanographic Data Center, Ocean Climate Laboratory, NODC Internal Report 17, Silver Spring MD, 17 p., 2002.

Day, J. J., Hawkins, E., and Tietsche, S.: Will Arctic sea ice thickness initialization improve seasonal forecast skill?, Geophys. Res. Lett., 41, 7566–7575, doi:10.1002/2014GL061694, 2014a.

Day, J. J., Tietsche, S., and Hawkins, E.: Pan-Arctic and Regional Sea Ice Predictability: Initialization Month Dependence, J. Climate, 27, 4371–4390, doi:10.1175/JCLI-D-13-00614.1, 2014b.

Day, J., Hawkins, E., and Tietsche, S.: Collection of Multi-model Data from the Arctic Predictability and Prediction On Seasonal-to-Interannual Time-scales (APPOSITE) Project, NCAS British Atmospheric Data Centre, doi:10.5285/45814db8-56cd-44f2-b3a4-92e41eaaff3f, 2015.

DelSole, T., Yan, X., Dirmeyer, P. A., Fennessy, M., and Altshuler, E.: Changes in seasonal predictability due to global warming, J. Climate, 27, 300–311, 2014.

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., Golaz, J.-C., Ginoux, P., Lin, S.-J., Schwarzkopf, M. D., Austin, J., Alaka, G., Cooke, W. F., Delworth, T. L., Freidenreich, S. M., Gordon, C. T., Griffies, S. M., Held, I. M., Hurlin, W. J., Klein, S. A., Knutson, T. R., Langenhorst, A. R., Lee, H.-C., Lin, Y., Magi, B. I., Malyshev, S. L., Milly, P. C. D., Naik, V., Nath, M. J., Pincus, R., Ploshay, J. J., Ramaswamy, V., Seman, C. J., Shevliakova, E., Sirutis, J. J., Stern, W. F., Stouffer, R. J., Wilson, R. J., Winton, M., Wittenberg, A. T., and Zeng, F.: The Dynamical Core, Physical Parameterizations, and Basic Simulation Characteristics of the Atmospheric Component AM3 of the GFDL Global Coupled Model CM3, J. Climate, 24, 3484–3519, doi:10.1175/2011JCLI3955.1, 2011.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., and Robinson, N.: Do seasonal to decadal climate predictions underestimate the predictability of the real world?, Geophys. Res. Lett., 41, 5620–5628, doi:10.1002/2014GL061146, 2014.

Eicken, H.: Ocean science: Arctic sea ice needs better forecasts, Nature, 497, 431–433, doi:10.1038/497431a, 2013.

Emmerson, C. and Lahn, G.: Arctic Opening: Opportunity and Risk in the High North, Tech. rep., Lloyds, Chattham House, 2012.

Flato, G. and Marotzke, Jochem and Abiodun, B. and Braconnot, P. and Chou, S. Chan and Collins, W. and Cox, P. and Driouech, F. and Emori, S. and Eyring, V. and others: in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 741–866, 2013.

Germe, A., Chevallier, M., Salas y Mélia, D., Sanchez-Gomez, E., and Cassou, C.: Interannual predictability of Arctic sea ice in a global climate model: regional contrasts and temporal evolution, Clim. Dynam., 43, 2519–2538, doi:10.1007/s00382-014-2071-2, 2014.

Goosse, H., Arzel, O., Bitz, C. M., de Montety, A., and Vancoppenolle, M.: Increased variability of the Arctic summer ice extent in a warmer climate, Geophys. Res. Lett., 36, L23702, doi:10.1029/2009GL040546, 2009.

Griffies, S. and Bryan, K.: A predictability study of simulated North Atlantic multidecadal variability, Clim. Dynam., 13, 459–487, 1997a.

Griffies, S. M. and Bryan, K.: Predictability of North Atlantic Multidecadal Climate Variability, Science, 275, 181–184, doi:10.1126/science.275.5297.181, 1997b.

Griffies, S. M., Winton, M., Donner, L. J., Horowitz, L. W., Downes, S. M., Farneti, R., Gnanadesikan, A., Hurlin, W. J., Lee, H.-C., Liang, Z., Palter, J. B., Samuels, B. L., Wittenberg, A. T., Wyman, B. L., Yin, J., and Zadeh, N.: The GFDL CM3 Coupled Climate Model: Characteristics of the Ocean and Sea Ice Simulations, J. Climate, 24, 3520–3544, doi:10.1175/2011JCLI3964.1, 2011.

Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J. J., Déqué, M., Doblas-Reyes, F. J., Fučkar, N. S., Germe, A., Hawkins, E., Keeley, S., Koenigk, T., Salas y Mélia, D., and Tietsche, S.: A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales: Arctic Sea-Ice Predictability and Prediction, Q. J. Roy. Meteorol. Soc., in press, doi:10.1002/qj.2401, 2014.

Hawkins, E., Tietsche, S., Day, J. J., Melia, N., Haines, K., and Keeley, S.: Aspects of designing and evaluating seasonal-to-interannual Arctic sea-ice prediction systems, Q. J. Roy. Meteorol. Soc., in press, doi:10.1002/qj.2643, 2015.

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., v. d. Hurk, B., v. Noije, T., v. d. Linden, E., and v. d. Wiel, K.: EC-Earth V2.2: description and validation of a new seamless earth system prediction model, Clim. Dynam., 39, 2611–2629, doi:10.1007/s00382-011-1228-5, 2012.

Holland, M. M., Bitz, C. M., Tremblay, B., and Bailey, D. A.: The role of natural versus forced change in future rapid summer Arctic ice loss, in: Arctic sea ice decline: observations, projections, mechanisms, and implications, edited by: DeWeaver, E., Bitz, C., and Tremblay, B., vol. 180 of Geophys. Monogr. Ser., AGU, Washington, 2008.

Holland, M. M., Bailey, D. A., and Vavrus, S.: Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3, Clim. Dynam., 36, 1239–1253, 2010.

Holland, M. M., Blanchard-Wrigglesworth, E., Kay, J., and Vavrus, S.: Initial-value predictability of Antarctic sea ice in the Community Climate System Model 3, Geophys. Res. Lett., 40, 1–4, doi:10.1002/grl.50410, 2013.

Jin, E. K., Kinter, J. L., Wang, B., Park, C.-K., Kang, I.-S., Kirtman, B. P., Kug, J.-S., Kumar, A., Luo, J.-J., Schemm, J., Shukla, J., and Yamagata, T.: Current status of ENSO prediction skill in coupled ocean–atmosphere models, Clim. Dynam., 31, 647–664, doi:10.1007/s00382-008-0397-3, 2008.

Johns, T. C., Durman, C. F., Banks, H. T., Roberts, M. J., McLaren, A. J., Ridley, J. K., Senior, C. A., Williams, K. D., Jones, A., Rickard, G. J., Cusack, S., Ingram, W. J., Crucifix, M., Sexton, D. M. H., Joshi, M. M., Dong, B.-W., Spencer, H., Hill, R. S. R., Gregory, J. M., Keen, A. B., Pardaens, A. K., Lowe, J. A., Bodas-Salcedo, A., Stark, S., and Searl, Y.: The new Hadley Centre climate model (HadGEM1): Evaluation of coupled simulations, J. Climate, 19, 1327–1353, 2006.

Jung, T., Kasper, M. A., Semmler, T., and Serrar, S.: Arctic influence on subseasonal midlatitude prediction, Geophys. Res. Lett., 41, doi:10.1002/2014GL059961, 2014.

Jung, T. Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., Dawson, J., Doblas-Reyes, F., Fairall, C, Goessling, H. F., Holland, M., Inoue, J., Iversen, T., Klebe, S., Lemke, P., Losch, M., Makshtas, A., Mills, B., Nurmi, P., Perovich, D., Reid, P., Renfrew, I. A., Smith, G., Svensson, G., Tolstykh, M., Yang, Q.: Advancing polar prediction capabilities on daily to seasonal time scales, BAMS, in press, doi:10.1175/BAMS-D-14-00246.1, 2016.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and von Storch, J. S.: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model, J. Adv. Model. Earth Syst., 5, 422–446, doi:10.1002/jame.20023, 2013.

Koenigk, T. and Mikolajewicz, U.: Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model, Clim. Dynam., 32, 783–798, doi:10.1007/s00382-008-0419-1, 2009.

Massonnet, F., Fichefet, T., and Goosse, H.: Prospects for improved seasonal Arctic sea ice predictions from multivariate data assimilation, Ocean Model., 88, 16–25, doi:10.1016/j.ocemod.2014.12.013, 2015.

Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., Ajayamohan, R. S., Fyfe, J. C., Tang, Y., and Polavarapu, S.: The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization, Mon. Weather Rev., 141, 2910–2945, doi:10.1175/MWR-D-12-00216.1, 2013.

Msadek, R., Vecchi, G. A., Winton, M., and Gudgel, R. G.: Importance of initial conditions in seasonal predictions of Arctic sea ice extent, Geophys. Res. Lett., 41, doi:10.1002/2014GL060799, 2014.

Notz, D., Haumann, F. A., Haak, H., Jungclaus, J. H., and Marotzke, J.: Arctic sea-ice evolution as modeled by Max Planck Institute for Meteorology's Earth system model, J. Adv. Model. Earth Syst., 5, 173–194, doi:10.1002/jame.20016, 2013.

Peterson, K. A., Arribas, A., Hewitt, H. T., Keen, A. B., Lea, D. J., and McLaren, A. J.: Assessing the forecast skill of Arctic sea ice extent in the GloSea4 seasonal prediction system, Clim. Dynam., doi:10.1007/s00382-014-2190-9, 2014.

Pohlmann, H., Botzet, M., Latif, M., Roesch, A., Wild, M., and Tschuck, P.: Estimating the decadal predictability of a coupled AOGCM, J. Climate, 17, 4463–4472, 2004.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res., 108, 4407, doi:10.1029/2002JD002670, 2003.

Schröder, D., Feltham, D. L., Flocco, D., and Tsamados, M.: September Arctic sea-ice minimum predicted by spring melt-pond fraction, Nature Clim. Change, 4, 353–357, doi:10.1038/nclimate2203, 2014.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R.: Uncertainty in modeled Arctic sea ice volume, J. Geophys. Res., 116, C00D06, doi:10.1029/2011JC007084, 2011.

Serreze, M. C., Strove, J.: Arctic sea ice trends, variability and implications for seasonal ice forecasting, Phil. Trans. R. Soc. A., 373:20140159, doi:10.1098/rsta.2014.0159, 2015.

Shaffrey, L. C., Stevens, I., Norton, W. A., Roberts, M. J., Vidale, P. L., Harle, J. D., Jrrar, A., Stevens, D. P., Woodage, M. J., Demory, M. E., Donners, J., Clark, D. B., Clayton, A., Cole, J. W., Wilson, S. S., Connolley, W. M., Davies, T. M., Iwi, A. M., Johns, T. C., King, J. C., New, A. L., Slingo, J. M., Slingo, A., Steenman-Clark, L., and Martin, G. M.: U.K. HiGEM: The New U.K. High-Resolution Global Environment Model–Model Description and Basic Evaluation, J. Climate, 22, 1861–1896, doi:10.1175/2008JCLI2508.1, 2009.

Shi, W., Schaller, N., MacLeod, D., Palmer, T. N., and Weisheimer, A.: Impact of hindcast length on estimates of seasonal climate predictability, Geophys. Res. Lett., 42, 1554–1559, doi:10.1002/2014GL062829, 2015.

Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Goessling, H. F., Handorf, D., Harig, S., Hiller, W., Juricke, S., Losch, M., Schröter, J., Sein, D. V., and Wang, Q.: Towards multi-resolution global climate modeling with ECHAM6–FESOM. Part I: model formulation and mean climate, Clim. Dynam., 44, 757–780, doi:10.1007/s00382-014-2290-6, 2014.

Sigmond, M., Fyfe, J. C., Flato, G. M., Kharin, V. V., and Merryfield, W. J.: Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system, Geophys. Res. Lett., 40, 529–534, doi:10.1002/grl.50129, 2013.

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., and Murphy, J. M.: Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model, Science, 317, 796–799, doi:10.1126/science.1139540, 2007.

Stephenson, S. R., Smith, L. C., Brigham, L. W., and Agnew, J. A.: Projected 21st-century changes to Arctic marine access, Clim. Change, 118, 885–899, doi:10.1007/s10584-012-0685-0, 2013.

Tietsche, S., Notz, D., Jungclaus, J. H., and Marotzke, J.: Predictability of large interannual Arctic sea-ice anomalies, Clim. Dynam., 41, 2511–2526, doi:10.1007/s00382-013-1698-8, 2013.

Tietsche, S., Day, J. J., Guemas, V., Hurlin, W. J., E. Keeley, S. P., Matei, D., Msadek, R., Collins, M., and Hawkins, E.: Seasonal to interannual Arctic sea ice predictability in current global climate models, Geophys. Res. Lett., 41, 1035–1043, doi:10.1002/2013GL058755, 2014.

Tietsche, S., Hawkins, E., and Day, J. J.: Atmospheric and oceanic contributions to irreducible forecast uncertainty of Arctic surface climate, J. Climate, 29, 331–346, doi:10.1175/JCLI-D-15-0421.1, 2016.

Wang, W., Chen, M., and Kumar, A.: Seasonal Prediction of Arctic Sea Ice Extent from a Coupled Dynamical Forecast System, Mon. Weather Rev., 141, 1375–1394, doi:10.1175/MWR-D-12-00057.1, 2013.

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T., Chikira, M., Ogura, T., Sekiguchi, M., Takata, K., Yamazaki, D., Yokohata, T., Nozawa, T., Hasumi, H., Tatebe, H., and Kimoto, M.: Improved Climate Simulation by MIROC5: Mean States, Variability, and Climate Sensitivity, J. Climate, 23, 6312–6335, doi:10.1175/2010JCLI3679.1, 2010.

Table 1. Details of simulations submitted to the APPOSITE database.

Model	CTRL length	Forcing year	Start dates	Start months	Ensemble size	References
HadGEM1.2	249	1990	10	Jan, May, Jul	16	Johns et al. (2006) Shaffrey et al. (2009)
MPI-ESM	200	2005	12 (Jul), 16 (Nov)	Jul, Nov	9 (Jul), 16 (Nov)	Notz et al. (2013) Jungclaus et al. (2013)
GFDL-CM3	200	1990	8	Jan, Jul	16	Donner et al. (2011) Griffies et al. (2011)
EC-Earth2.2	200	2005	9	Jul	8	Hazeleger et al. (2012)
MIROC5.2	100	2000	8	Jan, Jul	8	updated from Watanabe et al. (2010)
E6F	200	1990	18	Jan, Jul	9	Sidorenko et al. (2014)
CanCM4	45	transient (1970–2014)	32	Jan, Jul,	10	Sigmond et al. (2013) Merryfield et al. (2013)

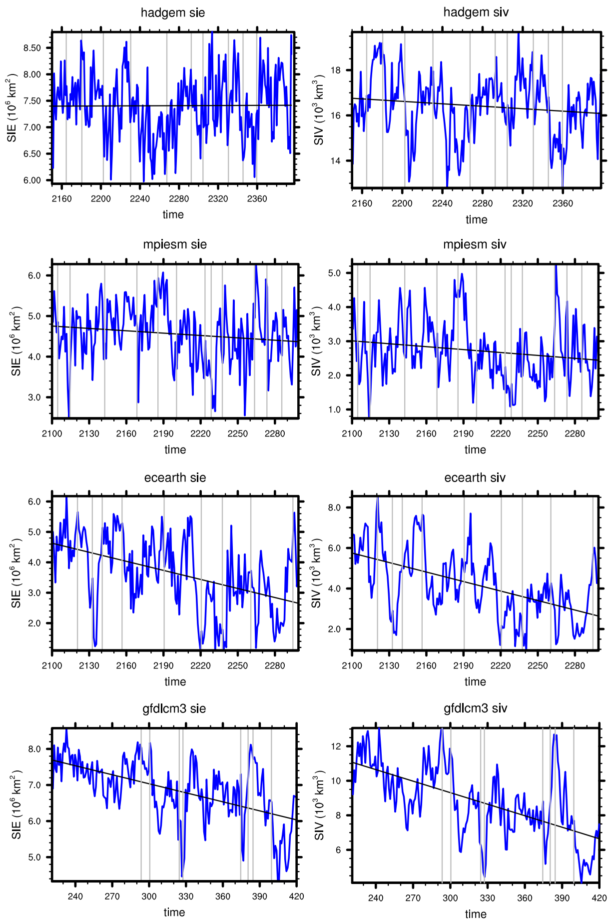
hadgem sie
SIE (10^6 km^2)
time
hadgem siv
SIV (10^3 km^3)
time
mpiesm sie
SIE (10^6 km^2)
time
mpiesm siv
SIV (10^3 km^3)
time
ecearth sie
SIE (10^6 km^2)
time
ecearth siv
SIV (10^3 km^3)
time
gfdlcm3 sie
SIE (10^6 km^2)
time
gfdlcm3 siv
SIV (10^3 km^3)
time

Figure 1.

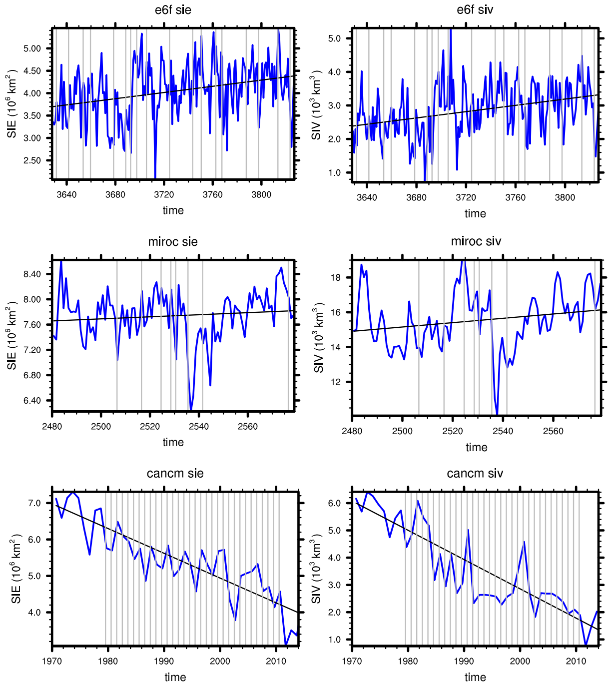

Figure 1. Timeseries of monthly mean September sea ice extent (sie, left column) and sea ice volume (siv, right column) in each model's control simulation (blue) with the line of best fit to data (black). Vertical grey lines indicate start years used to initialise simulations. Values on the time axis are model clock times, and do not correspond to the actual run-length of the simulation.

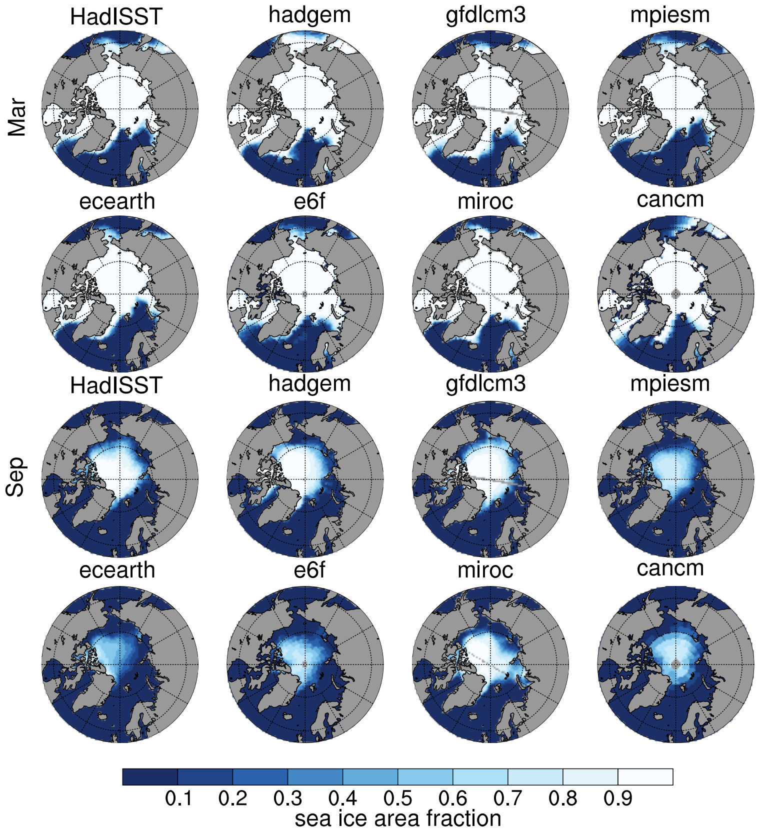

Figure 2. Average sea-ice concentration in present-day model control simulations and from HadISST (1983–2012) (Rayner et al., 2003).

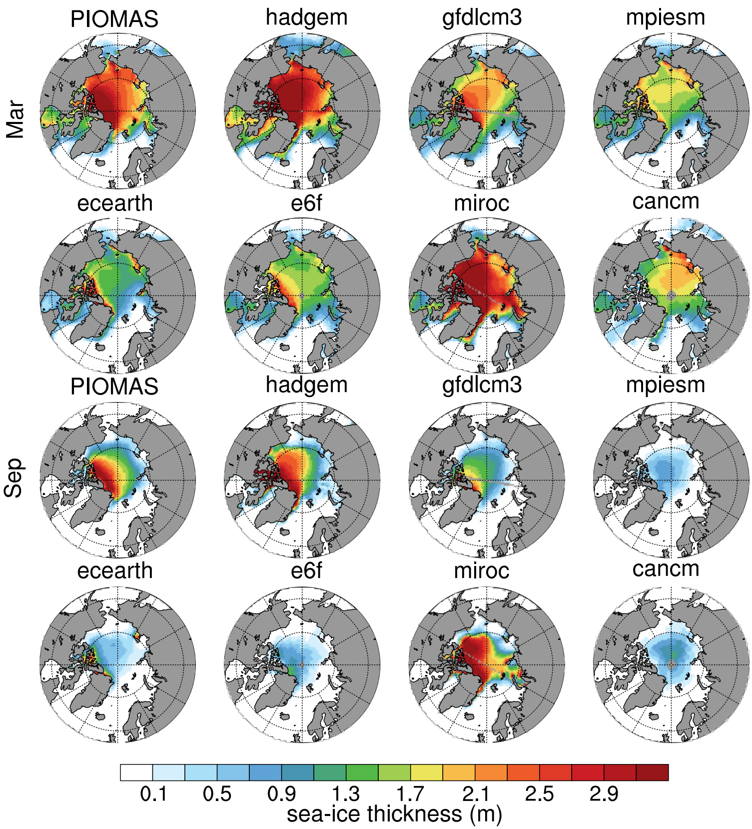

Figure 3. Average sea-ice thickness in present-day model control simulations and from PIOMAS (1983–2012) (Schweiger et al., 2011).

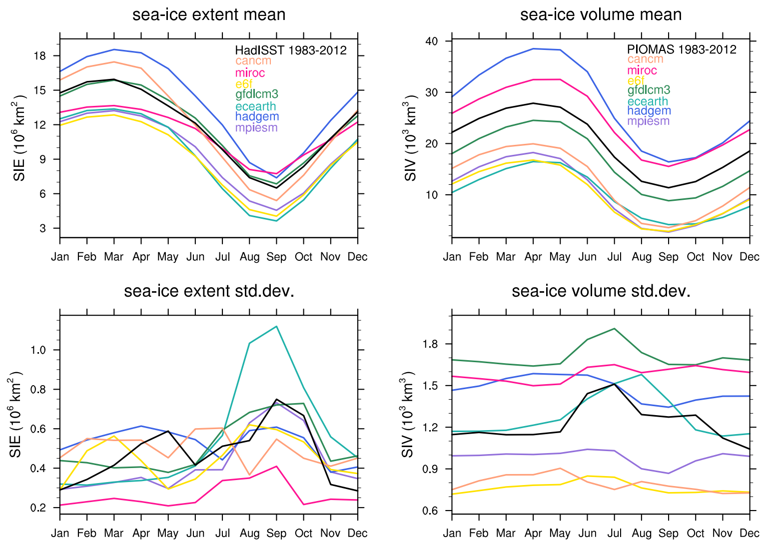

Figure 4. Seasonal cycle of monthly mean sea-ice extent **(a)**, volume **(b)** and standard deviation of sea ice extent **(c)** and volume **(d)** in present-day model control simulations. The HadISST observations of sea ice extent and PIOMAS reconstruction of ice volume are included as a reference. These data were linearly detrended prior to calculating the variance.

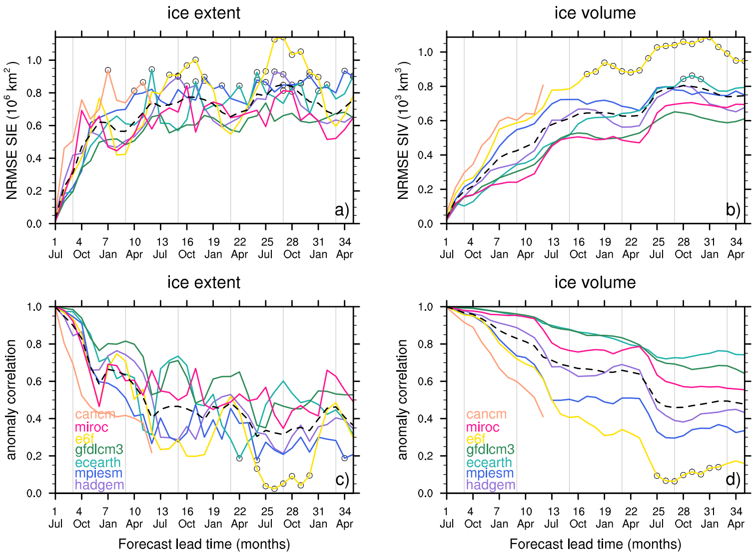

Figure 5. (a) and **(b)** Lead-time dependence of SIE NRMSE and SIV NRMSE for all models. **(c)** and **(d)** Lead-time dependence of SIE ACC and SIV ACC for all models. September and March are marked by thin gray vertical lines. Dashed lines represent the averages across models. Circles indicate where metrics do not indicate significant predictability (at 95 %). Updated from Tietsche et al. (2014).

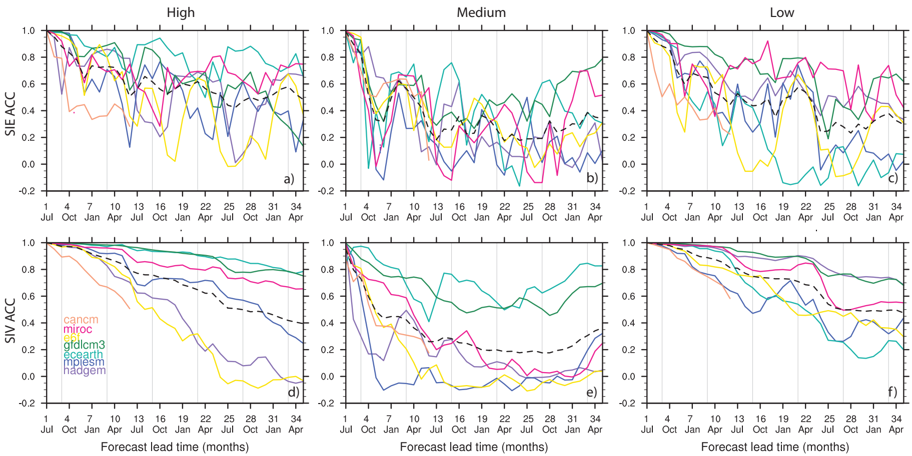

Figure 6. Top row: Anomaly correlation os sea ice extent, but calculated only for start dates with anomalously low, medium or high sea ice extent, relative to the control run climate. Bottom row, as top row but for sea ice volume, binned by volume.