

Supplementary Information: Towards seasonal Arctic shipping route predictions

N. Melia^{1,2}, K. Haines¹, E. Hawkins², and J. J. Day²

[1] National Centre for Earth Observation, Department of Meteorology, University of Reading, Reading, United Kingdom

[2] NCAS-Climate, Department of Meteorology, University of Reading, Reading, United Kingdom

Correspondence to: N. Melia (nathanael.melia@reading.ac.uk)

1 Skill metrics

The Brier Score (BS) and Brier Skill Score (BSS) can be used to ascertain forecast skill when the output of an ensemble forecast is binary, such as an open or closed sea route.

We start by defining our sea route opening event forecast for a particular year. Let $I(j) = [I_1(j), \dots, I_n(j)]$ be an n -member ensemble forecast of the binary open/closed events for the j^{th} day of the year, indicating whether the sea route was forecast to be open (=1) or closed (=0) by each member.

The control simulation contains only one deterministic realisation, the ‘truth’; the binary openness on day j is then denoted, $I_c(j)$.

From the initialised ensemble forecast, the forecast probability, $p_f(j)$, i.e. the probability of openness for the j^{th} day is the mean over the number of ensemble members e :

$$p_f(j) = \frac{1}{e} \sum_{i=1}^e I_i(j) \quad (1)$$

The Brier Score of the forecast, BS_f , for the whole length of the shipping season d days, is calculated as:

$$BS_f = \frac{1}{d} \sum_{j=1}^d \left(p_f(j) - I_c(j) \right)^2 \quad (2)$$

This means that BS_f measures how different the ensemble prediction is from the truth over the season (which here is July 1st – December 31st).

However to fully quantify the skill the average climatological conditions need to be taken into account. For example if the deterministic solution has a route open and the ensemble is 100% open

26 at that time this may appear as skilful, however, if the climatology from the entire control simulation
27 is also 100% open at this time this prediction will have no real additional skill¹.

28 The average climatological openness is calculated as the average for the j^{th} date, for all years, y , in
29 the control simulation:

$$p_c(j) = \frac{1}{y} \sum_{j=1}^y I_c(j) \quad (3)$$

30 The Brier Score of the climatology, BS_c , over the length of shipping season days, d , is calculated as:

$$BS_c = \frac{1}{d} \sum_{j=1}^d (p_c(j) - I_c(j))^2 \quad (4)$$

31 This means that BS_c measures the temporal variation of total openness from the deterministic
32 simulation compared to the climatology through the shipping season.

33 Finally, the Brier Skill Score, BSS , is a proper measure of skill, calculated by taking the forecast Brier
34 Score BS_f , and combining with the expected climatological Brier Score BS_c :

$$BSS = 1 - \frac{BS_f}{BS_c} \quad (5)$$

35
36

¹ This effect is seen for PC6 vessels - the apparent skill is lower than for OW vessel despite the July and even November forecasts being completely accurate for several months of the year. This is because the control simulation is 100% open for a few months, hence correctly predicting this period is attributed zero skill.

2 Mean sea ice thickness and variability

Figures S1 and S2 illustrate the higher variability in sea ice thickness (SIT) for July (representing sea route opening months) than for November (representing sea route closing months), particularly along the NSR.

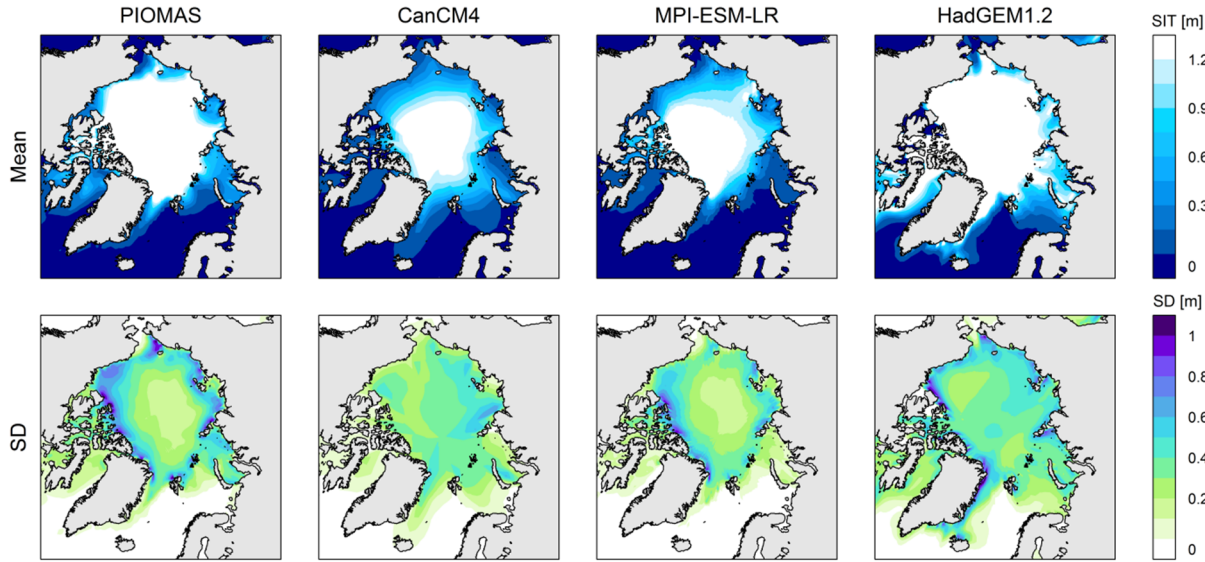


Figure S1. Mean July sea ice thickness and variability as Figure 1.

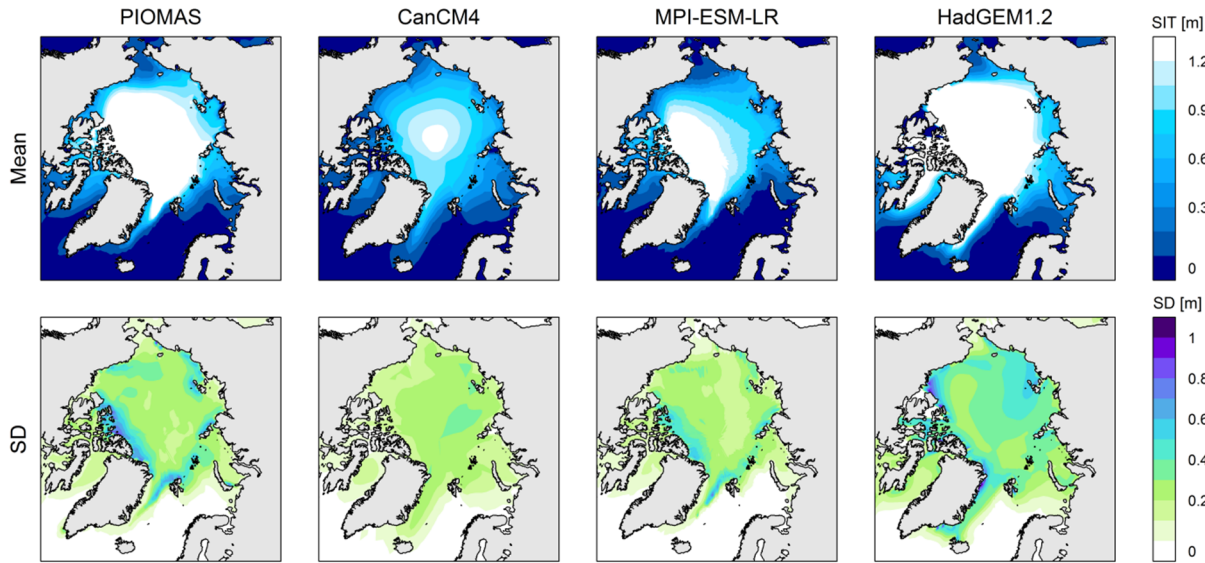


Figure S2. Mean November sea ice thickness and variability as Figure 1.

3 Climatological sea ice, SITCLIM initialisation

The July SITCLIM simulations in Section 6 exhibit a mean BSS = -0.45, less than both January SITCLIM (BSS = 0.18) simulations, and climatological skill (BSS = 0, by definition). This section will investigate why the July SITCLIM forecasts with climatological SIT appear to perform worse than climatology

when initialised close to the validation date, and why this mechanism does not appear to have hindered the January SITCLIM forecasts.

3.1 Jul SITCLIM bias

The first clue to investigate the source of the July SITCLIM error is to ascertain if the forecasts show a bias. The July SITCLIM results presented in Figure 5 (main manuscript) appear to be more open than the other simulations, particularly for closed routes.

The mean of all years from Figure 5 is shown in Figure S3. The July SITCLIM all-years—ensemble-mean has a $\approx 20\%$ open bias September to November. This explains the poor July SITCLIM performance compared to January.

Figure S4 illustrates the locations where SITCLIM has led to changes in SIT that lasts until at least the middle of the shipping season. The abundance of positive SIT anomalies (green) in the anomaly plots shows that for these closed years (as prescribed by the deterministic simulation) the replacement with climatological SIT has led to a low SIT (open route) bias. This is also evident in the orange contour representing the PC6 Class Vessel SIT threshold in Figure S4; comparing the SITCLIM to the SITINIT plots it is clear the prohibitively thick sea ice around the Russian islands in the SITINIT plots have been replaced by lower SIT in the SITCLIM simulations.

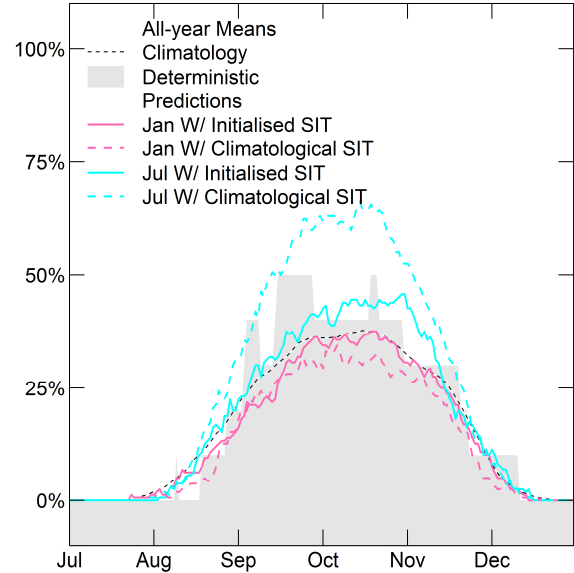


Figure S3. Mean HadGEM1.2 perfect model NSR forecast openings for PC6 vessels for the January and July SITINIT and SITCLIM simulations.

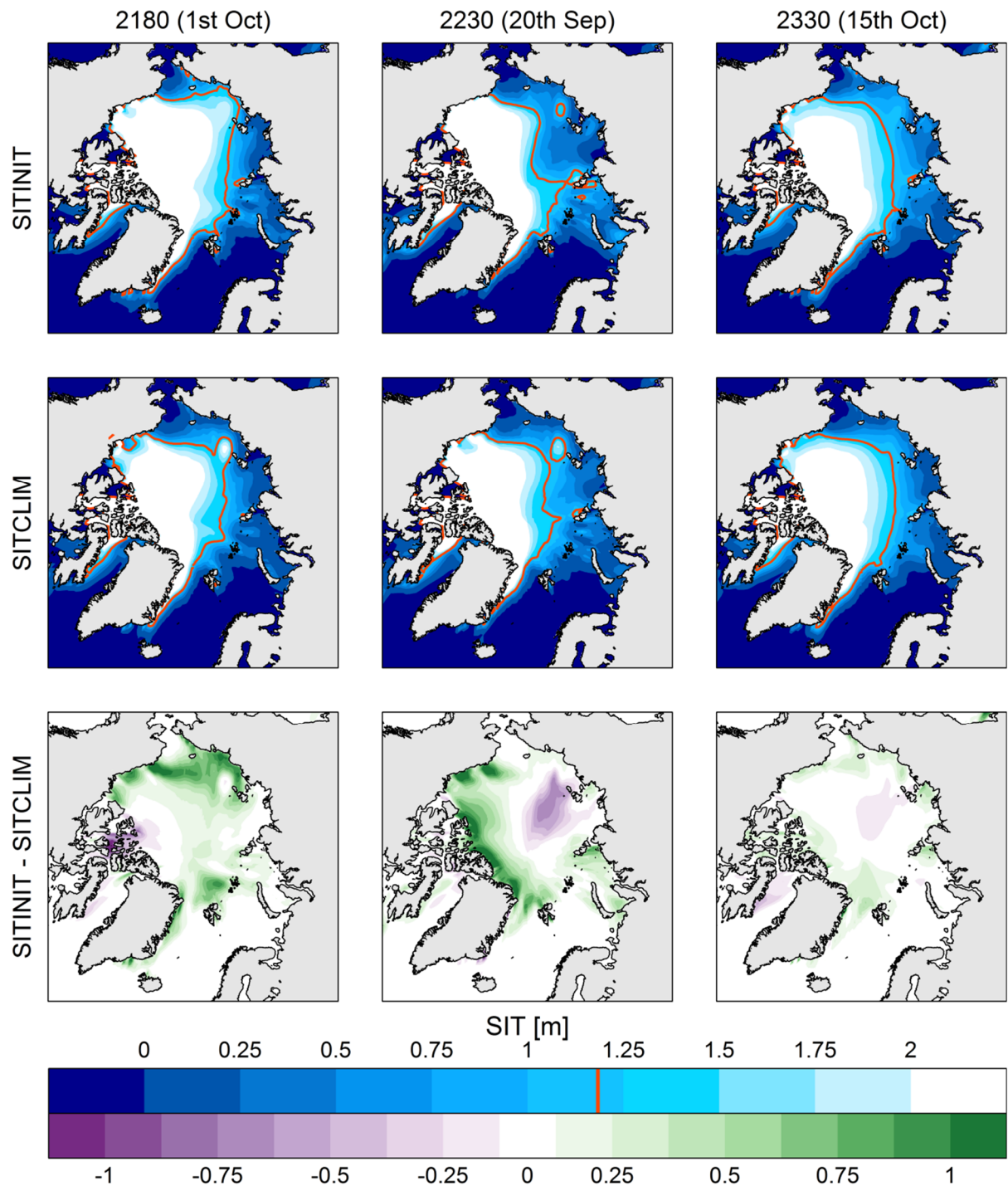


Figure S4. HadGEM1.2 forecast ensemble-mean SIT for three July 1st initialisations which exhibit closed routes in the deterministic control simulation. The dates shown for each year were selected to illustrate when the difference between SITINIT and SITCLIM would be largest. The orange contour indicates the limit of the PC6 vessels accessibility. The bottom row anomaly plots show the difference at the validation dates between the SITINIT and SITCLIM simulations.

We have established that the implementation of SITCLIM conditions initialised in forecasts from July can lead to a relatively long-lived negative SIT anomaly for high background SIT conditions. The fact that this effect occurs in high SIT years gives a hint as to the mechanism: for relatively high SIT years

the climatological SIT will be lower, hence the implementation of the SITCLIM process will remove SIT. From July through summer the prevailing conditions are melting, hence there is no significant mechanism that can recover the induced negative SIT anomaly and so these conditions can persist until following freeze season. Simply the positive sea ice feedbacks during the melt season maintain or even exacerbate the induced SITCLIM anomalies.

During the January initialisations SITCLIM conditions may recover to the SITINIT conditions as negative sea ice feedbacks dominate during the freeze season, manifested by the SIT — sea ice growth rate relationship, whereby thicker ice grows slower than thin ice, so in the growing season anomalies recover back to the SITINIT sea ice evolution.

To investigate this behaviour we examine the evolution of SIT over the NSR from the January and July initialisations comparing SITINIT with SITCLIM.

3.2 July SITCLIM SIT evolution

In the July SITCLIM simulations large negative SIT anomalies may persist until winter as during the summer the primary SIT growth mechanism is absent. The simulated years 2180 and 2292 (Figure S5) show this phenomenon whereby the negative anomaly introduced by SITCLIM is maintained until the following spring. In 2180 this SITCLIM anomaly is maintained and in 2292 the anomaly grows, likely due the positive feedbacks. For the high ice years 2230 and 2330 the SITCLIM anomaly is smaller so the evolution of SITCLIM is similar to SITINIT; despite this these years also give negative BSS. Figure S4 shows that this is because of removal of SIT by SITCLIM at key choke points has resulted in an open NSR for some ensemble members, with no chance of this existing SITINIT thick sea ice regions replenishing during the summer. It is these lasting effects present in the high ice years which accounts for the negative BSS scores, this was confirmed by removing the high ice years and calculating the mean BSS of the remaining years: July SITINIT = 0.1, July SITCLIM = 0.0.

In low SIT years (Figure S5: 2202, 2267, and 2395), where SITCLIM has higher SIT than SITINIT, SITCLIM may also exhibit poor BSS if the anomaly is large enough despite melting conditions. This is the case with year 2359 where the SITCLIM anomaly never fully recovers to SITINIT. In 2267 the anomaly is smaller, however SITCLIM still takes three months (until October) to recover, whereby the majority of the season is already over. In 2202 the SITCLIM anomaly is sufficiently small that the evolution is similar to SITINIT and hence good BSS.

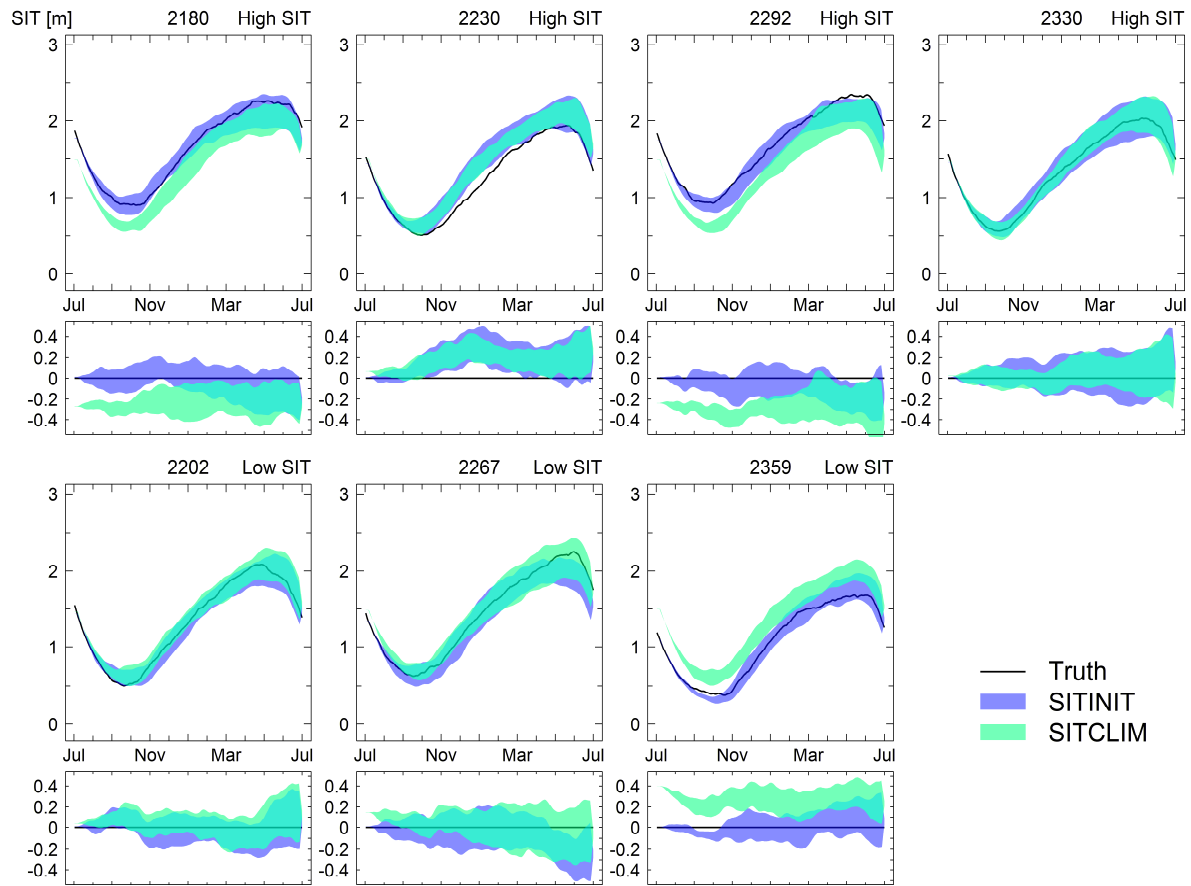


Figure S5. HadGEM1.2 July initialised sea ice thickness evolution averaged over the NSR for, ‘High SIT’ years — upper two rows, and ‘Low SIT’ years — lower two rows; absolute SIT evolution in square plots above and SIT anomalies with respect to the ‘Truth’ (the years control simulation) below. The SITINIT and SITCLIM swaths represent the ensemble forecast envelope (range).

3.3 January initialisation

On average January SITCLIM skill is better the July SITCLIM skill. The primary reason for this is likely due to the time it takes for the SITCLIM anomalies to recover to the SITINIT evolution; for the July initialisations there is not enough time for the larger anomalies to recover but from January there is for more cases. A complication when comparing the SITINIT and SITCLIM simulations from the January initialisations is that at these greater lead times the ‘Truth’ may fall out of either SITINIT and SITCLIM ensemble envelope. This is the case with ‘High SIT’ years 2230 and 2330; however the BSS for these are good even though the forecast envelope has clearly overestimated SIT. This is because the BSS is based on an event and does not penalise for forecasting significantly over the event threshold. The January SITCLIM ensembles will also depict a larger range than the July SITCLIM so can achieve some apparent skill from this.

For the remaining ‘High SIT’ years 2180 and 2292 the SITCLIM anomalies are large but have somewhat recovered for the open season. They do however remain with a distinct low bias compared to SITINIT, that is one the open threshold resulting in low skill (BSS = 0.22 and 0.26). For the ‘Low SIT’ years 2202 and 2267 the SITCLIM anomalies have all had time to recover by August, the

2359 anomaly decreased under negative feedbacks until May when the anomaly still present started to increase, presumably due to the melt seasons positive feedbacks.

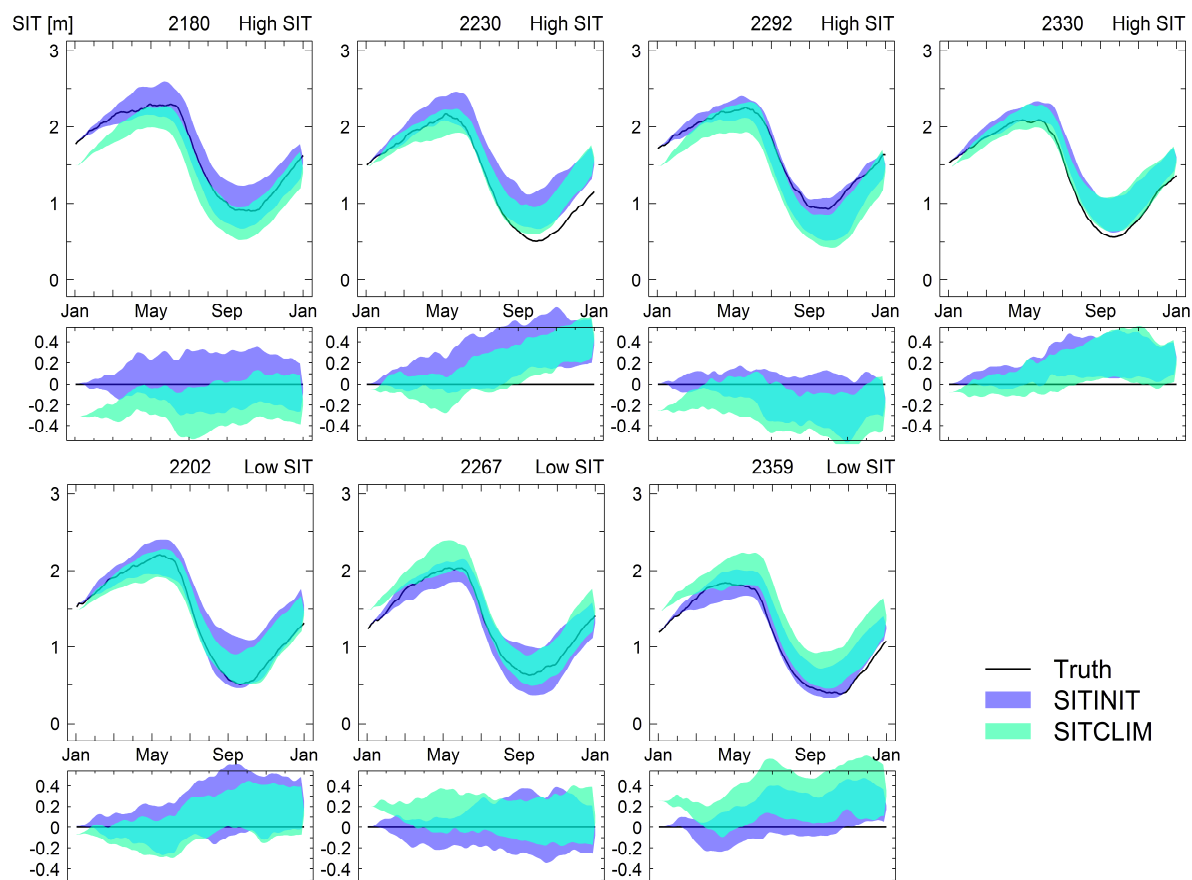


Figure S6. HadGEM1.2 July initialised sea ice thickness evolution averaged over the NSR for, ‘High SIT’ years — upper two rows, and ‘Low SIT’ years — lower two rows; absolute SIT evolution in square plots above and SIT anomalies with respect to the ‘Truth’ (the years control simulation) below. The SITINIT and SITCLIM swaths represent the ensemble forecast envelope (range).

The general reduction of SIT anomalies in the SITCLIM January simulations during the freeze season is consistent with negative feedbacks in this environment and provides a mechanism whereby the lack of SIT information in January may not substantially degrade January forecast skill for specific ice states. This concurs with [Guemas *et al.*, 2016] who find that the memory for predictions being held by the ocean heat content rather than SIT for forecasts initialised during the freeze season.

4 Discussion

The potential role of positive and negative feedbacks to initialised anomalies have been presented here. However a definitive investigation of these feedback mechanisms would need to assess the surface energy balance which is beyond the scope of this paper.

These mechanisms are state dependant; to robustly attribute the differing behaviours in the different state would require more initialisations. The HadGEM1.2 experiments benefited from 16 ensemble members, however only 10 start dates. Given that skill is state dependant this is in fact

150 more restricted to four high, three medium and three low years; ideally there would be at least 10 of
151 each to reduce the signal to noise ratio and enable robust conclusions to be made.

152 **References**

153 Guemas, V., M. Chevallier, M. Déqué, O. Bellprat, and F. Doblas-Reyes (2016), Impact of sea ice
154 initialization on sea ice and atmosphere prediction skill on seasonal timescales, *Geophys. Res. Lett.*,
155 43(8), 3889-3896, doi: 10.1002/2015gl066626.

156