Revision of "Seasonal transition dates can reveal biases in Arctic sea ice simulations"

We thank the two referees and the editor for their constructive and positive feedback on the discussion paper. Below, we have included descriptions of changes made due to the detection of two coding bugs in the process of making revisions (which do not change any conclusions), followed by point-by-point replies to the two reviews.

Description of two corrections made in the revised manuscript

1. Satellite-era medians versus means
In the initial submission of the manuscript, a metric called the "satellite-era median" was defined to represent a measure-of-center of the average spatial distribution of each seasonal transition date from 1979-2014. This process was described in the Methods section:

> "To quantitatively compare the models to each other and to observations in a pan-Arctic sense, we take the median of the resulting distribution, shown as a histogram in Fig. 2. This value is referred to as the "satellite-era median".

When addressing the comments of the referees, code bugs were found associated with the medians, so that what was described as medians in the original submission were actually means, and the means of the satellite data and the means for models with more than 30 ensemble members were calculated from histograms of different bin sizes than the CMIP models with less members. We have analyzed the impacts of these code bugs, and have found that they do not change any of the conclusions.

Using the means is not more or less valid than using the medians, as both still achieve the stated original goal: they represent a valid measure-of-center that accounts for differences in spatial coverage. However when using mean values, it is more accurate to use area-weighted spatial means, since means from the distributions lose decimal places in the binning process (otherwise they are equivalent). **As the conclusions are not impacted and both are valid metrics, we have chosen to continue to show the means (now calculated as area-weighted spatial means for accuracy) and changed the Methods to accurately reflect this, which reduced the changes compared to the discussion paper.**

For reference, details of the effects of using area-weighted spatial means (referred to from here on as "means") versus medians from the areal distributions (referred to from here on as "medians") are described below, based on the detailed analysis we have performed. **We have also edited the Methods section to briefly describe how using medians instead of means affects the results.**

Detailed description of the impact of using means instead of medians:
Due to the skew of the distributions (Figure R1, which was included as Figure 2 in the original submission and is now Figure S3 in the Supplement in response to Referee #2), using mean values results in earlier spring transition dates and later fall transition dates compared to using median values. In the models, the average difference between the mean and median for all transition dates is 7 days and the maximum difference is 19 days (associated with freeze onset). In the satellite data, the average difference is also 7 days and the maximum difference is 16 days (also associated with freeze onset). To demonstrate an example of these differences, Tables R1 and R2 below show the median and mean melt onset and freeze onset days for each model and the number of days between them.

None of the conclusions are affected by the use of the mean versus the median. However, if we were to use the medians, two findings would be less-pronounced than reported based on the means.

1) **Melt onset in the CNRM models is later than in other models and satellite data:** The difference between the melt onset in the CNRM models and other models and satellite observations is smaller when medians are compared (see Table R1). However, as all models except the CNRM models and the CESM LE still fall earlier than or on the same day as the satellite data when medians instead of means are used, this does not affect the overall validity of this finding.

2) Internal variability cannot explain the later simulated freeze onset and freeze-up: Using medians, many models still demonstrate later freeze onset and freeze-up dates than observed, but fewer models fall outside of the maximum range of internal variability compared to the satellite data (Table R3). In turn, differences between modeled and observed melt seasons and open water periods are smaller as well. However, this does not affect our interpretation of the results since the models still tend to fall later than the satellite data in terms of freeze onset and freeze-up and longer in terms of melt season and open water period--the differences are just less pronounced and may more often be due to internal variability.

The other results in the manuscript are even less affected by using medians instead of means. The spreads between ensemble members in models with more than 30 members vary by an average of two days. The scatter plots presented in Figure 10 still maintain the same patterns, and the average differences in correlation coefficients for Tables 3 and 4 are of magnitude 0.01.

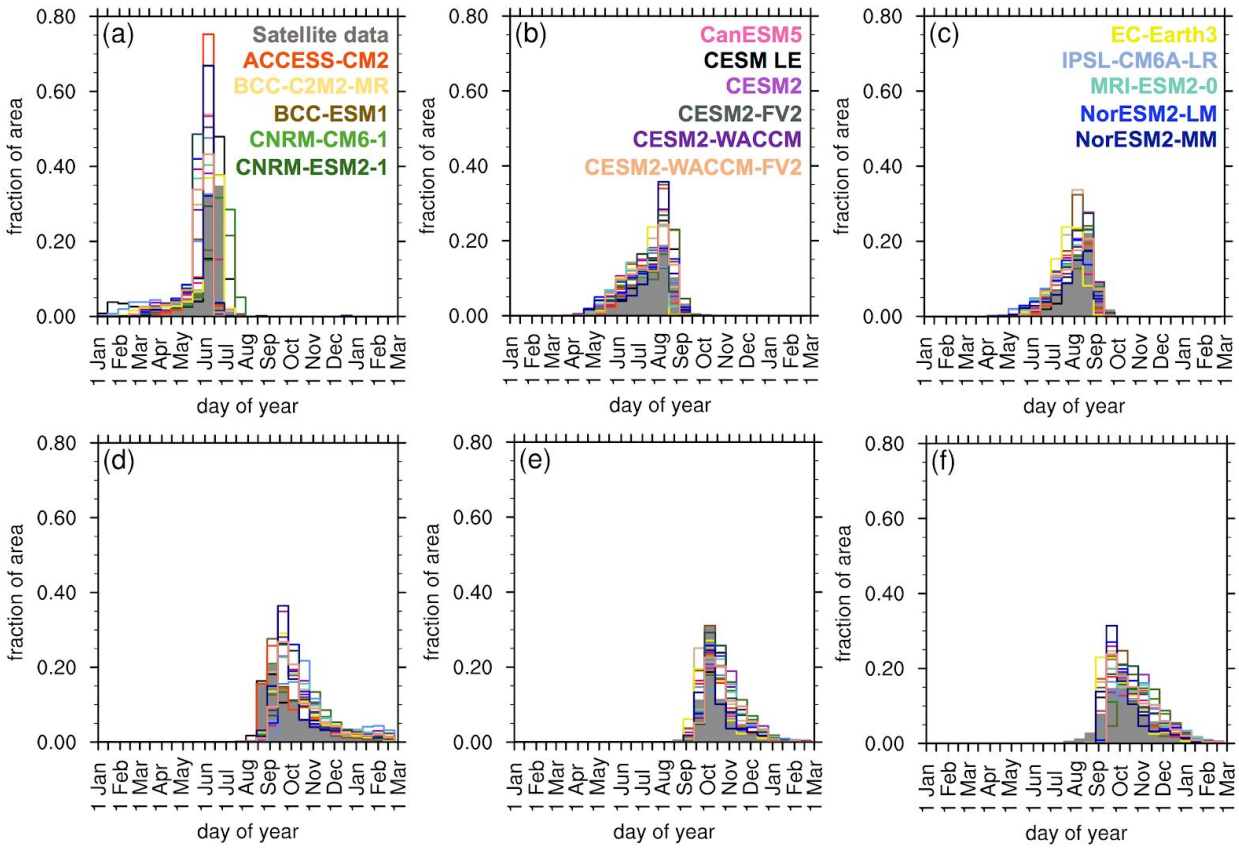

Figure R1. Area distributions of the average of each metric from 1979-2014: (a) melt onset (b) opening (c) break-up (d) freeze onset (e) freeze-up and (f) closing. Metrics are averaged from 66-84.5 N for satellite data (filled gray) and the first ensemble member of each model (all other colors). All models and satellite data are represented in each panel (a)-(f), but the color labels are distributed across panels (a)-(c).

	Mean melt onset date	Median melt onset date	Difference in days (mean minus median)
ACCESS-CM2	3-Jun	5-Jun	-2
BCC-CSM2-MR	24-May	29-May	-5
BCC-ESM1	30-May	2-Jun	-3
CanESM5	3-Jun	6-Jun	-3
CESM2	20-May	28-May	-8
CESM2-FV2	22-May	27-May	-5
CESM2-WACCM	23-May	29-May	-6
CESM2-WACCM-FV2	21-May	28-May	-7
CNRM-ESM2-1	14-Jun	18-Jun	-4
CNRM-CM6-1	18-Jun	21-Jun	-3
EC-Earth3	2-Jun	10-Jun	-8
IPSL-CM6A-LR	15-May	30-May	-15
MRI-ESM2-0	22-May	27-May	-5
NorESM2-LM	21-May	26-May	-5
NorESM2-MM	29-May	3-Jun	-5
CESM LE	29-May	16-Jun	-18
Satellite data	6-Jun	10-Jun	-4

Table R1. Pan-Arctic, satellite-era (1979–2014) mean melt onset dates, median melt onset dates and the difference between them in days for the first ensemble member of each model and satellite data. This table has been adapted from Table 3 in the original manuscript.

	Mean freeze onset date	Median freeze onset date	Difference in days (mean minus median)
ACCESS-CM2	6-Oct	17-Sep	19
BCC-CSM2-MR	8-Oct	27-Sep	11
BCC-ESM1	2-Oct	22-Sep	10
CanESM5	16-Oct	29-Sep	17
CESM2	23-Oct	10-Oct	13
CESM2-FV2	18-Oct	3-Oct	15
CESM2-WACCM	17-Oct	2-Oct	15
CESM2-WACCM-FV2	17-Oct	2-Oct	15
CNRM-ESM2-1	28-Oct	18-Oct	10

CNRM-CM6-1	21-Oct	8-Oct	13
EC-Earth3	10-Oct	28-Sep	12
IPSL-CM6A-LR	3-Nov	20-Oct	14
MRI-ESM2-0	25-Oct	11-Oct	14
NorESM2-LM	17-Oct	3-Oct	14
NorESM2-MM	8-Oct	22-Sep	16
CESM LE	7-Oct	24-Sep	13
Satellite data	4-Oct	18-Sep	16

Table R2. Pan-Arctic, satellite-era (1979–2014) mean freeze onset dates, median freeze onset dates and the difference between them in days for the first ensemble member of each model and satellite data. This table has been adapted from Table 3 in the original manuscript.

	Freeze onset		**Freeze-up**	
	Number of models later than satellite data	Number of models outside greatest estimated range of internal variability	Number of models later than satellite data	Number of models outside greatest estimated range of internal variability
Means	15	10	11	5
Medians	15	5	13	2

Table R3. The number of models out of sixteen that show later freeze onset and freeze-up dates in terms of their medians and means. This table also shows the number of models out of sixteen that are later than the satellite data by at least the largest number of days estimated to represent internal variability.

2. Correction in Supplementary Table S4

The correlation coefficients in the columns for freeze onset, freeze-up and closing in Supplementary Table S4 were meant to portray the relationship between the fall transition metrics and the March ice area of the following year, but by accident the correlation coefficients shown were for the same year (instead of the following year). This does not impact the conclusions, as differences in correlation coefficients for the third through sixth columns of Table S4 are of magnitude 0.01. The values in Table S4 have now been corrected.

Point by point reply to Review comments:
Anonymous Referee #1

In this study, the authors evaluate climate model performance for describing the Arctic sea ice seasonal cycle with a series of metrics that describe stages of the melt and freeze-up periods. Modeled sea ice concentrations and surface temperatures are used to approximate significant dates of the melt and freeze cycle obtained from passive microwave satellite observations. The authors find that the models capture a similar asymmetry in the melt/freeze cycle where the melting period is longer than the corresponding freeze-up period as seen in observations. Overall, the models generally agree with observations on the timing of spring melt, but several show delays in the timing of the freeze-up period relative to the observations. The differences between models for these variables exceed expected differences due to internal variability of the model, thus, the authors propose using the seasonal sea ice transition metrics to evaluate climate model performance. I find that the paper is very well written, interesting, and thorough in reporting the results of the study. I have a few very minor comments that the authors should address before publication as outlined below.

> We thank the referee very much for their constructive comments. We have made all the suggested changes (details below).

Minor Comments

1. L 205-206: I'm not sure exactly what you mean by the model spread shifting "earlier toward the satellite data". Can you please expand or rephrase this in the text?

> To clarify this, we have edited this sentence. It now reads: "Excluding the CNRM models (which show particularly late mean melt onset dates and are explored further in Sect. 4.5), the model spread (May 15--June 3) shifts earlier, and the mean melt onset dates from the remaining models all occur earlier than the satellite data."

2. L 206: Inflow regions are not specifically defined anywhere in the paper. It would be worth it to state where these are (e.g., Bering Strait, etc.).

> As requested, we have now added a definition of the "inflow regions" as the Chukchi Sea, Barents Sea and Greenland Sea, and of the "Atlantic inflow regions" as the Barents Sea and Greenland Sea.

3. L 249: It would be beneficial to expand a bit on the ice concentration metric used by Markus et al. (2009) when the algorithm does not detect a clear freeze onset signal from the brightness temperatures. Specifically, that the threshold used by Markus is 80% SIC, which in theory makes some unknown quantity of the satellite freeze onset dates more comparable to the closing dates, than freeze-up dates. This is likely contributing to the instances where the freeze transition dates are out of order (e.g. as you state in lines 247-248).

> As suggested, we have now expanded this discussion of the back-up sea ice concentration metric used by Markus et al. (2009): "In satellite data, simultaneous freeze onset and freeze-up dates may in part be explained by the satellite retrieval algorithms: the PMW retrieval algorithm for freeze onset uses an 80% ice concentration metric to derive freeze onset at locations where the date can not be reliably derived using the weighted brightness temperature scheme (Markus et al., 2009). This would skew the freeze

onset dates later and make them more similar to the closing dates. Hence, the use of ice concentration by both the freeze onset and freeze-up retrieval algorithms may contribute to cases where the dates are not sequential. A detailed assessment of this is not possible, however, as the data does not contain information on how often this back-up method is employed."

4. Figures 3-8: Since Jan – Mar are repeated twice in the color scale, it would be easier for readers to see that the repeated dates in the blue colors are for the following year if this was denoted on the scale markings somehow.

We agree and have adjusted Figures 2-7 such that the two Januarys are labeled with "year" and "year+1". We have also bolded the line on the color bar denoting $\text{January}_{\text{year+1}}$.

Technical Corrections

1. L 30: Typo – Melt ponds _decrease_ the albedo of the surface

Corrected.

2. L 328: Typo – the former _through_ the formation of...

Corrected.

Anonymous Referee #2

The manuscript suggests to evaluate CMIP6 sea ice simulations using ice seasonality metrics. The authors define a set of seasonal metrics based on sea ice concentration and surface temperature simulated by CMIP6 and CESM earth system models. A recent satellite database describing several ice seasonality metrics is also used to evaluate models.

A first part is focused on transition period between the different melt and freezing dates. They find an asymmetry between spring ice loss and fall ice growth in both satellite observations and model simulations. They also show that most models have a late freeze onset compared to observations.

In the second part the correlation between seasonal metrics and sea ice area and thickness is studied. The authors find a good correlation between freezing dates , sea ice area and thickness. These correlations allow to highlight sea ice biases in some models which are compensated by other processes. The authors give the example of CNRM model which has biases in melting and in sea ice thickness but which represent realistic sea ice area for the wrong reasons.

I found this paper very interesting. It includes newly observation database and suggests an interesting and novel approach to evaluate earth system models sea ice simulations. I stress the good work the authors made by analysing a large set of earth system models with several members besides satellite observations. I think that this work can make a great contribution to the literature after following comments are addressed in the context of a minor revision.

> We thank the referee very much for their constructive comments. We have made all recommended changes, including adding the suggested new Figures and Table.

Specific Comments

1. P1 line 13 "the spread between climate model projections of sea ice has been on the order of millions of square kilometers in Coupled Model Intercomparison Project (CMIP)" . Can you specify about which parameter you are talking? Sea ice coverage ?

> As suggested, we have changed the sentence to clarify that we are talking about sea ice area here. The new sentence reads: "the spread between climate model projections of sea ice area has been on the order of millions of square kilometers in Coupled Model Intercomparison Project (CMIP)".

2. P4 line 92 "(select figures using all available members are provided in the Supplement)." Can you specify which figures ?

> As requested, we have added which figures this refers to: “(a version of Figure 8 using all available ensemble members is provided as Figure S2 in the Supplement).”

3. P9 to P17 : I think it would be beneficial to add maps of the four intra-seasonal periods (melt , freeze , seasonal loss-of-ice and seasonal gain-of-ice periods) and a table of spatial median (and standard deviation) for each model and observation as for table 4. Moreover, you look at the difference between the spatial of the

metrics to describe the median of the intra-seasonal periods. But as median(A) - median(B) /= med(A-B), calculating the intra-seasonal period for each pixel before doing the spatial median seems more appropriate.

As suggested, a table of the mean values and model spreads of the intra-seasonal periods has been added as Table 4 in the manuscript. Spatial plots for both intra-seasonal periods and inter-seasonal periods have been provided in the Supplementary (Figues S4-S10). Modeled standard deviations generally agree with satellite data standard deviations. This has been noted in text and the standard deviations have been provided in Supplementary Tables S3 and S4.

With respect to the differences, we have added text to the manuscript to clarify this process: "As with the transition dates, the inter-seasonal and intra-seasonal periods are calculated at each grid cell before taking the area-weighted spatial means."

4. P11 lines 205-207 "the model spread (May 15- June 3) " : Is it really June 3 or is it June 13 here?

Excluding the CNRM models, the latest satellite-era median melt onset date occurs on June 3 (ACCESS-CM2 and CanESM5). We have edited the text so that this is clearer.

5. P14 line 228 : What do you mean by "internal variability of the satellite data" ?

We have rearranged this sentence to say "The maximum range in mean freeze onset dates due to internal variability is 11 days (Table 3) and the majority of the model means (ten out of sixteen) are more than 11 days later than the satellite data, indicating that this delay of the mean freeze onset in the models is not only due to internal variability." The original sentence with the uncorrected values was, "Only five of the sixteen models fall within the maximum range of internal variability (10 days) of the satellite data (Table 3)."

6. p20 line 302 " (Supplementary Table S3)" : I guess you mean Table S4.

The supplementary table numbers have changed due to the addition of new tables, and in Specific Comment #13 below, we address how we have added text to clarify differences between the tables to avoid confusion. In the original submission, this did refer to Table S3.

7. p22 line 320 "This lack of relationship is a strong indication that the spatial coverage of break-up dates is not sufficient for describing pan-Arctic sea ice feedbacks. " : I wonder if the lack of relationship between March mean ice thickness and break up date can be explained by the inverse relation between ice growth and thickness which explains that the thinner the ice, the more efficient the growth. This relation should temper the delay in break up (see Bitz & Roe, 2004 and Lebrun et al. 2019).

Bitz, C. M., Holland, M. M., Hunke, E. C. and Moritz, R. E.: Maintenance of the Sea-IceEdge, J. Climate, 18(15), 2903?2921, doi:10.1175/JCLI3428.1, 2005.

Lebrun, M., Vancoppenolle, M., Madec, G. and Massonnet, F.: Arctic sea-ice-free season projected to extend into autumn, The Cryosphere, 13(1), 79-96, doi:https://doi.org/10.5194/tc-13-79-2019, 2019.

As suggested, we have evaluated the relationship between March mean ice thickness and break-up in the context of the listed papers, and we have added text related to the break-up dates, as well as text related to the freeze-up dates, where this also applies:

LINE 345: “Increases in ice thickness after March may dampen the relationship between thin March ice and an earlier break-up date, since some models show faster ice growth from March to April in areas of thin March ice rather than thick March ice (supporting past work on ice growth rates (Bitz and Roe, 2005)). However, this pattern is not seen in all models and thus cannot fully account for the weakness of the relationships between March ice thickness and break-up.”

LINE 330: “With respect to the other fall transition metrics, we find that statistically significant correlations between March ice thickness and freeze-up/closing (which are both based on ice concentration) are less consistent between models, and generally stronger for the closing dates rather than freeze-up dates (Table 7). Other relationships involving freeze-up and spring sea ice of the following year, such as the relationship between the timing of freeze-up and the next year's break-up, have been shown to be dampened by the tendency of thin ice to grow faster than thicker ice (Bitz and Roe, 2005, Lebrun et al., 2019). The growth rate of thin ice, in addition to the spatial coverage of the freeze-up dates, may be limiting the impact that a late freeze-up date has in reducing the following year's March mean ice thickness.”

8. p22 line 247 " indicating that the impact of seasonal transition biases can be be large" you should remove a "be"

Corrected.

9. Figure 2 : This figure seems not describe in detail in the main text. You should move it in supplementary.

As suggested, we have moved Figure 2 to the Supplement (now Figure S3).

10. Figures 3 and 6: Can you remind the definition criteria for melt and freeze onset dates in both caption as you did for fig 4,5,7 and 8 ?

As suggested, we have added the phrase “(defined using surface temperature in the models and brightness temperatures in the satellite data)” to the captions of these figures (now Figures 2 and 5) to clarify how the melt and freeze onset dates are derived.

11. Table 3 : What do you mean by "spread" ?

Here the word “spread” is referring to the difference between the earliest and the latest dates found between the first member of all models (the all-model spread) and using the first thirty members of each model (the models marked with *). To clarify this, we have now defined “spread” in the Methods section and in the captions of Tables 3-5 that refer to spread.

12. Table 3 - Table 4 : Can you also add a spatial standard deviation for each model and observation?

As discussed above, modeled standard deviations generally agree with satellite data standard deviations. This has been noted in text and the standard deviations have been provided in Supplementary Tables S3 and S4.

13. Table 6 or Table S4 : Caption for both tables are exactly the same. I guess it is a mistake you should fix.

We have added text to clarify what is different for figures and tables with very similar captions. For example in Table S6 (table numbers have changed due to the addition of new tables) we have added, “As in Table 7, correlation coefficients (R-values) between seasonal sea ice transition dates but with March sea ice area instead of sea ice thickness from 1979–2014”.

Seasonal transition dates can reveal biases in Arctic sea ice simulations

Abigail Smith[1], Alexandra Jahn[1], and Muyin Wang[2,3]

[1]Department of Atmospheric and Oceanic Sciences and Institute of Arctic and Alpine Research, University of Colorado Boulder
[2]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington
[3]Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration

Correspondence: Abigail Smith (abigail.l.smith@colorado.edu)

Abstract. Arctic sea ice experiences a dramatic annual cycle, and seasonal ice loss and growth can be characterized by various metrics: melt onset, break-up, opening, freeze onset, freeze-up and closing. By evaluating a range of seasonal sea ice metrics, CMIP6 sea ice simulations can be evaluated in more detail than by using traditional metrics alone, such as sea ice area. We show that models capture the observed asymmetry in seasonal sea ice transitions, with spring ice loss taking about ~~1.5–2~~ 1-2 months longer than fall ice growth. The largest impacts of internal variability are seen in the inflow regions ~~of~~ for melt and freeze onset dates, but all metrics show pan-Arctic model spreads exceeding the internal variability range, indicating the contribution of model differences. Through climate model evaluation in the context of both observations and internal variability, we show that biases in seasonal transition dates can compensate for other unrealistic aspects of simulated sea ice. In some models, this leads to September sea ice areas in agreement with observations for the wrong reasons.

1 Introduction

Metrics of seasonality have been under-utilized in evaluating sea ice in climate models, due to a lack of long-term observational products, the required daily model output and the complexities in defining seasonal Arctic sea ice transitions. However, new process-based metrics for model evaluation are much needed—the spread between climate model projections of sea ice area has been on the order of millions of square kilometers in Coupled Model Intercomparison Project (CMIP) Phases 3, 5, and 6 (Stroeve et al., 2012; SIMIP-Community, 2020), while the causes of the model spread remain largely unknown. Furthermore, the sources of model biases can be obscured by models that show realistic sea ice areas for the wrong reasons. Seasonal sea ice transitions can provide additional process-based metrics to assess climate models. Newly available observational data (Steele et al., 2019) and model output from CMIP6 models (Notz et al., 2016) allow such model assessment for the first time. In this study, we assess how different metrics of seasonal sea ice transitions are represented in models and observations, and evaluate how these metrics can inform our understanding of simulated Arctic sea ice throughout the year. To do this, we utilize observations and sixteen global climate models, including three sets of ensembles with at least 30 members. Using this rich data set, we evaluate model biases in the context of both the observed sea ice state and multiple simulated representations of internal variability.

2 Background: Seasonal transitions in the Arctic sea ice cover

Arctic sea ice exhibits a large annual cycle, with a difference of approximately 8 million square kilometers between the maximum area reached in March and the minimum area in September. From spring to fall, the sea ice experiences various stages of transition forced by both the atmosphere and the ocean (Steele et al., 2010; Persson, 2012; Ballinger et al., 2019). In the spring, clouds formed by northward warm air advection trap downwelling longwave radiation, initiating melt on the surface of the sea ice or in the snowpack on top of it (Persson, 2012; Ballinger et al., 2019). As liquid water collects on the snow and sea ice, it forms melt ponds. Melt ponds ~~increase~~ decrease the albedo of the surface: snow-covered ice has an albedo of 0.85, while the albedo of melt ponds ranges between 0.1-0.5 (Perovich et al., 2002). Shortwave absorption causes thermodynamic ice loss, and regional studies show that top melt dominates during the early summer (Steele et al., 2010). As the ice breaks up, larger areas of open ocean facilitate greater solar absorption (the albedo of open water is 0.07 (Pegau and Paulson, 2001)) and ice divergence. Energy is absorbed by the surface ocean (Timmermans, 2015) and as solar heating declines in the late summer, ice melt becomes dominated by bottom melt (Steele et al., 2010). After the annual sea ice minimum in September, ice growth begins. Congelation ice growth along existing ice generally begins before frazil ice growth in the open ocean, meaning that areas where ice is retained throughout the summer experience earlier ice growth than areas of open water (Smith and Jahn, 2019). As fall progresses, the Arctic loses shortwave input. Temperatures decline and ice growth continues through the winter, reaching the maximum area in March.

One metric alone cannot capture the range of seasonal transitions seen in the Arctic, so individual transitions have been characterized by many different definitions in both satellite data and models. Seasonal transition metrics are often referred to interchangeably when they are in fact defined in very different ways. Pan-Arctic satellite retrievals of seasonal sea ice transitions are largely based on passive microwave brightness temperatures. Retrieval algorithms have been created to derive pan-Arctic seasonal sea ice metrics, such as melt onset and freeze onset, directly from brightness temperatures for the entire satellite era (Markus et al., 2009; Drobot and Anderson, 2001; Belchanksy et al., 2004; Bliss and Anderson, 2014; Bliss et al., 2017). Despite ~~ideal~~ great spatial and temporal coverage, melt and freeze onset dates are difficult to utilize for model evaluation. This is in part due to the variations between retrieval algorithms, which can introduce large differences in both magnitude and trends of observed melt onset dates (Bliss et al., 2017). Furthermore, brightness temperatures are not simulated in climate models, so model definitions of melt and freeze onset must be based on other simulated variables. There are multiple ~~valid~~ possible variables for diagnosing melt and freeze onset, such as surface temperature, thermodynamic ice growth, and snowmelt, and the choice of variable has been shown to impact which processes are captured by the dates, as well as their comparability to satellite data (Smith and Jahn, 2019).

Another strategy for defining seasonal sea ice transitions is to create metrics based on ice concentration, a variable that has equally good spatial and temporal satellite data coverage, since satellite-observed ice concentration is derived from passive microwave brightness temperatures (Comiso et al., 1997). While this introduces some error through sea ice concentration retrieval algorithms (Ivanova et al., 2015), seasonal sea ice metrics based on ice concentration provide more direct comparisons between models and observations than the current comparisons made between melt and freeze onset. Ice break-up, retreat,

freeze-up and advance have been defined using ice concentration data in satellite data (Stammerjohn et al., 2012; Serreze et al., 2016; Stroeve et al., 2016; Bliss et al., 2019) and in model studies (Barnhart et al., 2016; Wang et al., 2018). However, these studies are often difficult to compare directly, since the definitions themselves vary substantially in terms of the region and date range studied, the selected threshold of ice concentration and the criteria that the threshold must meet (e.g. last day greater than 15% vs. less than 15% two days in a row). In some cases, definitions are also created to fill specific user needs, such as seasonal navigation (Johnson and Eicken, 2016). A selection of previously used metrics defined using ice concentration are described ~~by~~ in Table S1 in the Supplement to provide an overview.

3 Data and Methods

In this study we use satellite data to evaluate the performance of fifteen CMIP6 models and the Community Earth System Model Large Ensemble (CESM LE) in terms of their seasonal sea ice transitions in the Arctic from 1979–2014. By utilizing model ensembles, we are able to account for the role of internal variability in modeling the seasonality of Arctic sea ice. As there is no single metric that fully describes seasonal sea ice changes, we utilize a variety of metrics that have been developed for both models and observations. All ~~spatial medians~~ seasonal transition date means and means of other sea ice variables are calculated between 66-84.5°N in both models and satellite data in order to exclude the largest polar hole in satellite data. We define "inflow regions" as the Chukchi Sea, Barents Sea and Greenland Sea, with the Barents and Greenland Seas referred to as "Atlantic inflow regions."

3.1 Global coupled climate models

CMIP establishes a set of common experiments for global climate model simulations to quantify how the Earth system responds to forcing, as well as to identify the sources and consequences of model biases (Eyring et al., 2016). This study uses models from the most recent phase, CMIP6, in order to evaluate the current state of sea ice simulation. Models are selected for analysis based on the availability of two daily sea ice variables: sea ice concentration (CMIP6 variable name: siconc) and the surface temperature of the sea ice or snow on sea ice (CMIP6 variable name: sitemptop). Our study utilizes all CMIP6 models that met this criteria by March 4th, 2020, which includes fifteen models from nine different institutions (ACCESS-CM2, BCC-CSM2-MR, BCC-ESM1, CanESM5, CESM2, CESM2-FV2, CESM2-WACCM, CESM2-WACCM-FV2, CNRM-ESM2-1, CNRM-CM6-1, EC-Earth3, IPSL-CM6A-LR, MRI-ESM2-0, NorESM2-LM, NorESM2-MM)~~(Table S2)~~. As the scope of the study is limited to the satellite era, we use the historical forcing experiment from each model for the period that overlaps with satellite data (1979-2014). All models are kept on their native grids to minimize errors related to interpolation and regridding. Ocean and ice model component details for all models are provided in Supplementary Table S2.

The number of available ensemble members varies by model, with some models providing as few as three members and others as many as thirty-five with the required daily variables. Here we use the first ensemble member (r1i1p1f1 or r1i1p1f2) from each model for inter-model comparisons and evaluation against satellite data. To assess the internal variability of the seasonal sea ice metrics, the two CMIP6 Models with at least 30 members (CanESM5 and IPSL-CM6A-LR, hereafter referred

to as IPSL) are utilized, in addition to the CESM LE. All of the coupled global models have a nominal ocean resolution of 1°. Relevant variables are available at a daily temporal resolution for 40 members in the CESM LE, 35 members in the CanESM5 and 30 members in the IPSL. When evaluating internal variability, we utilize the first 30 members from the CESM LE and CanESM5 for comparison to each other and IPSL in order to standardize the sample size. The results are insensitive to the subsetting of ensemble members (~~select figures~~ a version of Fig. 8 using all available ~~members are provided~~ ensemble members is provided as Fig. S2 in the Supplement).

In previous work, the CESM LE was employed to compare multiple model definitions of melt and freeze onset (Smith and Jahn, 2019). Hence, the CESM LE is utilized here to leverage what is already known about modeled seasonal sea ice transitions in evaluating CMIP6 models, even though the CESM1.1 used for the CESM LE is not a CMIP6 model and does not use CMIP6 forcing. Nonetheless, the CESM LE can be compared with the CMIP6 models over the period 1979-2014, as the CMIP5 RCP8.5 forcing is not substantially different from the CMIP6 historical forcing over the period 2006-2014 (O'Neill et al., 2016). Furthermore, the CESM LE is also a useful addition to the CMIP6 models because it adds diversity to the sea ice models used for evaluating internal variability: the CESM LE uses the CICE Version 4.0 sea ice model, while CanESM5 and IPSL both use the Louvain-la-Neuve Sea Ice Model Version (LIM Version 2.0 and LIM Version 3.0 respectively).

3.2 Satellite data

In order to evaluate the climate models against observations, we use the Arctic Sea Ice Seasonal Change and Melt/Freeze Climate Indicators from Satellite Data, Version 1 (Steele et al., 2019). This dataset includes seasonal sea ice indicators from March 1st, 1979 through February 27, 2017, derived from sea ice concentration data from the NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration and brightness temperature observations from the DMSP SSM/I-SSMIS Daily Polar Gridded Brightness Temperatures. Indicators (referred to here as seasonal sea ice transition metrics) are described in Sect. 3.3. Data are gridded to a 25 km resolution grid. We calculate the sea ice area from the NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, accessed through the Walsh et al. (2019) dataset.

3.3 Defining seasonal sea ice transitions

Establishing a set of metrics for studying the seasonality of Arctic sea ice is important for comparing models and observations, as well as interpreting the relationships between transition times and other sea ice characteristics. Here we utilize a range of seasonal sea ice transition dates and periods to study multiple thermodynamic phases of the ice that may be relevant to our physical understanding of the sea ice. These metrics are summarized in Tables 1 and 2.

3.3.1 Melt onset, freeze onset and the melt season

The melt season length is defined as the number of days between melt onset and freeze onset. The melt season length has been utilized as a parameter to investigate energy absorption of the Arctic surface ocean and relationships have been found between the melt season length and sea ice extent (Stroeve et al., 2014). The metrics of melt onset and freeze onset are used to describe

Dates	Date range	Timing	Variable	Threshold
Melt onset	1 Jan to 31 December	for 3 days	surface temperature	above -1°C
Freeze onset	29 June to 15 May	for 21 days	surface temperature	below -1.8°C
Break-up	1 March to SIC minimum date	last day	ice concentration	below 15%
Freeze-up	SIC minimum date to 28 February	first day	ice concentration	above 15%
Opening	1 March to SIC minimum date	last day	ice concentration	below 80%
Closing	SIC minimum date to 28 February	first day	ice concentration	above 80%

Table 1. Definitions of the seasonal transition dates, including the date range, timing criteria, variable and threshold used. Definitions based on ice concentration are designed to be comparable to Steele et al. (2019)

.

Intra-seasonal periods	# of days between
Melt period	Melt onset and opening
Freeze period	Freeze onset and freeze-up
Seasonal loss-of-ice period	Opening and break-up
Seasonal gain-of-ice period	Freeze-up and closing
Inter-seasonal periods	**# of days between**
Melt season	Melt onset and freeze onset
Open water period	Break-up and freeze-up
Outer ice-free period	Opening and closing

Table 2. Definitions of the periods of time between the seasonal transition dates, including shorter, intra-seasonal periods of transition as well as longer, inter-seasonal periods. The outer ice-free period and the seasonal loss-of-ice and gain-of-ice periods were defined in Steele et al. (2019).

the first date of continuous sea ice melt and freeze at each grid cell for each year. Melt and freeze onset are meant to capture a change of phase between water and ice. For melt onset, this means water on the surface of the ice or snowpack. For freeze onset, the change of phase refers to either congelation or frazil ice growth.

In satellite retrievals, continuous melt and freeze onset are defined using the brightness temperature of the surface because brightness temperature is sensitive to the phase of water (Markus et al., 2009; Steele et al., 2019). Brightness temperatures are collected at the 19V and 37V polarizations from SMMR,SSM/I and SSMIS sensors. Melt and freeze onset dates are derived from weighted brightness temperature parameters to determine early melt and freeze onset and continuous melt and freeze onset, and the retrieval algorithm (known as PMW) is described fully in Markus et al. (2009). In this study we use continuous melt and freeze onset because these dates are more representative of a seasonal transition in the sea ice compared to early melt onset. The AHRA dataset provides an alternative set of melt onset dates that are derived from passive microwave brightness

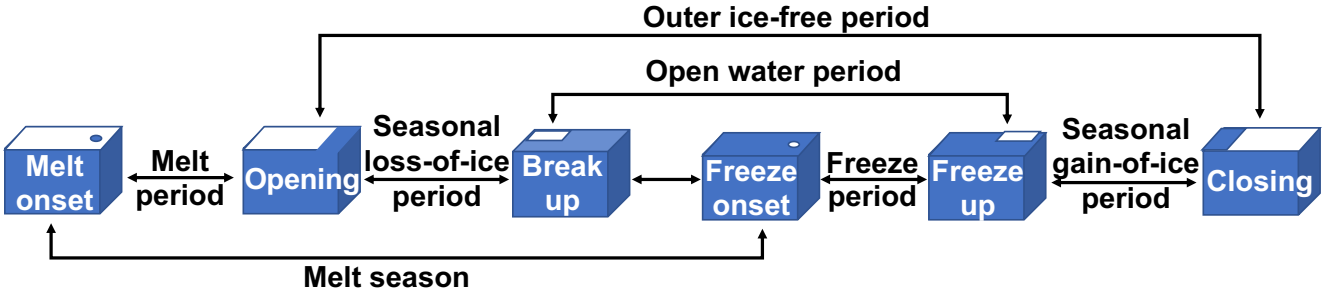

Figure 1. Conceptual diagram of seasonal sea ice transitions, beginning with spring melt onset and ending with fall ice closing. Transition dates (Table 1) as well as inter-seasonal and intra-seasonal transition periods (Table 2) are labeled.

temperatures using the AHRA retrieval algorithm instead of the PMW retrieval algorithm. However, the AHRA melt onset dates are more representative of early melt (Drobot and Anderson, 2001), so they are not utilized in this study.

Because climate models do not simulate brightness temperatures, another definition must be used to identify continuous melt and freeze onset dates within models. Although there is no single model definition that fully captures the processes represented by the brightness temperature-based satellite data, recent work demonstrates that melt and freeze onset dates derived from surface temperature are comparable, particularly when considering the range of internal variability (Smith and Jahn, 2019). We therefore utilize model definitions of melt and freeze onset developed in Smith and Jahn (2019), based on the surface temperature passing below/above a given threshold. For melt onset, a threshold of -1 °C is used to minimize the impacts of daily variability and maintain comparability with previous studies (Jahn et al., 2012; Mortin and Graversen, 2014). For freeze onset, a threshold equal to the freezing point of ocean water (-1.8 °C) is used.

3.3.2 Break-up, freeze-up and the open water period

The open water period, also known as the inner ice-free period (Bliss et al., 2019), is defined as the number of days between ice break-up and freeze-up (also commonly referred to as ice retreat and advance). The open water period has been utilized as a metric to study variability and trends in the sea ice (Serreze et al., 2016; Barnhart et al., 2016) and seasonal predictability of the ice (Stroeve et al., 2016).

Of the seasonal sea ice transition dates investigated here, the definitions of break-up, freeze-up and the open water period vary the most across the literature (Table S1). In the models, we use the definitions for break-up, freeze-up and the open water period used in Steele et al. (2019) to allow for comparison with observations. Of the definitions identified and described in Supplementary Table S1, the Steele et al. (2019) definitions are most similar to those established by Stroeve et al. (2016). Break-up is defined as the last day that sea ice concentration passes below the threshold of 15% between March 1 and the annual sea ice concentration minimum date (Bliss et al., 2019). Freeze-up is defined as the first day that sea ice concentration passes above the 15% threshold between the sea ice concentration minimum date and February 28 of the following year. The open water period is defined as the number of days between break-up and freeze-up.

3.3.3 Date of opening, date of closing and the outer ice-free period

The outer ice-free period has been used the least frequently as a metric of Arctic sea ice seasonality, and it is based on the dates of opening and closing defined by Steele et al. (2015). The Steele et al. (2015) definitions are applied in the Steele et al. (2019) dataset~~(Bliss et al., 2019)~~, as described in Bliss et al. (2019). The date of opening is defined as the first day that sea ice concentration passes below the threshold of 80% between March 1 and the annual sea ice concentration minimum date. Likewise, the date of closing is defined as the first day that sea ice concentration passes above the 80% threshold between the sea ice concentration minimum date and February 28 of the following year. By definition, opening must occur before break-up and freeze-up must occur before closing. However, dates of opening and closing are not limited solely by the existence of break-up and freeze-up dates: if ice concentration falls below/above 80% but not below/above 15%, there will still be an opening/closing date. This means that the areal coverage of opening/closing dates is generally larger than those of break-up/freeze-up dates.

3.3.4 Melt period and freeze period

In addition to the inter-seasonal periods (melt season, open water period and outer ice-free period) we describe four intra-seasonal periods between the established dates (melt period, seasonal loss-of-ice period, freeze period and seasonal gain-of-ice period) (Fig. 1). The melt period is designed to capture the rate of transition between snow and sea ice to the initial appearance of open water, and it is defined as the number of days between sea ice melt onset and the date of opening. Similarly, the freeze period is defined as the number of days between freeze onset and freeze-up, and is designed to describe the ~~rate of~~ length of the transition between initial ice growth ~~water and when~~ and the time at which an area stops being "ice-free" by exceeding the 15% concentration threshold.

3.3.5 Seasonal loss-of-ice period and seasonal gain-of-ice period

The seasonal loss-of-ice period and the seasonal gain-of-ice period were established in ~~Steele et al. (2019); Bliss et al. (2019)~~Steele et al. (2 Bliss et al. (2019). The seasonal loss-of-ice period is defined as the number of days between sea ice opening and break-up, and the seasonal gain-of-ice period is defined as the number of days between freeze-up and closing (Bliss et al., 2019). The seasonal loss-of-ice period and the seasonal gain-of-ice period can only be calculated at grid cells where both of their respective dates exist for that year (e.g. both a date of opening and break-up are needed for a valid seasonal loss-of-ice period). The seasonal loss-of-ice period describes how quickly the ice concentration transitions from 80% to 15%, while the seasonal gain-of-ice period describes the rate of transition between 15% and 80% ice concentration.

3.3.6 Accounting for differences in spatial coverage

Variability between models depends on the selected metric for evaluating seasonal sea ice changes. Over the satellite era, opening, break-up, freeze-up and closing each have an ice concentration boundary ~~,~~ where there are no existing dates beyond that boundary, because the ice concentration does not pass the chosen threshold. The models have different sea ice areas (Supplementary Fig. S1) and the position of the ice concentration boundary varies substantially between them. It is therefore

important to compare the models between each other and the satellite data in a way that captures these differences in spatial coverage. Using one ensemble member from each CMIP6 model and one ensemble member from the CESM LE, we find the mean of each characteristic at each grid cell over 1979–2014. We then ~~find the spatial distribution of~~ take the spatial mean by weighing each grid cell value ~~versus the fractional area that the value takes up north of~~ by its respective area between 66–84.5°N. This ~~allows for a representation of the~~ value is referred to as the "satellite-era mean" and represents the pan-Arctic nature of each characteristic~~without taking a pan-Arctic mean, which would obscure many spatial differences. To quantitatively compare the models to each other and to observations in a pan-Arctic sense, we take the median of the resulting distribution, shown as a histogram in Fig. S3. This value is referred to as~~ . Supplementary Fig. S3 shows the value of each transition date versus the percent area it spans, demonstrating that the spatial distribution of the modeled dates and their skews are realistic compared to satellite data. While some model differences are less pronounced when using satellite-era medians instead of satellite-era means, the ~~"~~results are generally insensitive to the chosen measure-of-center. As with the transition dates, the inter-seasonal and intra-seasonal periods are calculated at each grid cell before taking the area-weighted spatial means.

For each seasonal transition metric, model spreads in satellite-era ~~median"~~means are defined as the number of days between the earliest and latest simulated date. Spreads are calculated between the first ensemble member of all models as well as between the first thirty ensemble members of CESM LE, CanESM5 and IPSL. Figures 2–7 show each of the seasonal sea ice metrics derived from satellite data, one ensemble member from each available CMIP6 model and one ensemble member of the CESM LE averaged over 1979–2014. Each figure includes stippling to show where the characteristic exists for less than 20% of years in the time period.

~~Area distributions of the average of each metric from 1979–2014: (a) melt onset (b) opening (c) break-up (d) freeze onset (e) freeze-up and (f) closing. Metrics are averaged from 66-84.5N for satellite data (filled gray) and the first ensemble member of each model (all other colors). All models and satellite data are represented in each panel (a)-(f), but the color labels are distributed across panels (a)-(c).~~

4 Results

Results are presented in five sections. In Sect. 4.1–4.3 we describe the pan-Arctic observed and simulated seasonal sea ice transition metrics from 1979-2014. In Sect. 4.4 and 4.5 we compare observed and simulated relationships between the various seasonal sea ice transition metrics and sea ice area and thickness.

4.1 Spring transitions

We find that the transition from sea ice melt onset to break-up takes two to three months in both satellite data and models. Satellite data show that melt onset generally occurs between April and June over most regions of the Arctic (Fig. 2a), with the ~~median~~ mean date of melt onset occurring on ~~May 30. The median~~ June 6. The mean date of opening (July ~~8~~16) occurs about 40 days after melt onset and the ~~median~~ mean break-up date (~~July 28) occurs 20~~ August 4) occurs 19 days after opening (Table 3). This indicates that the most time-consuming aspect of the observed spring ice loss is the transition between ~~water~~

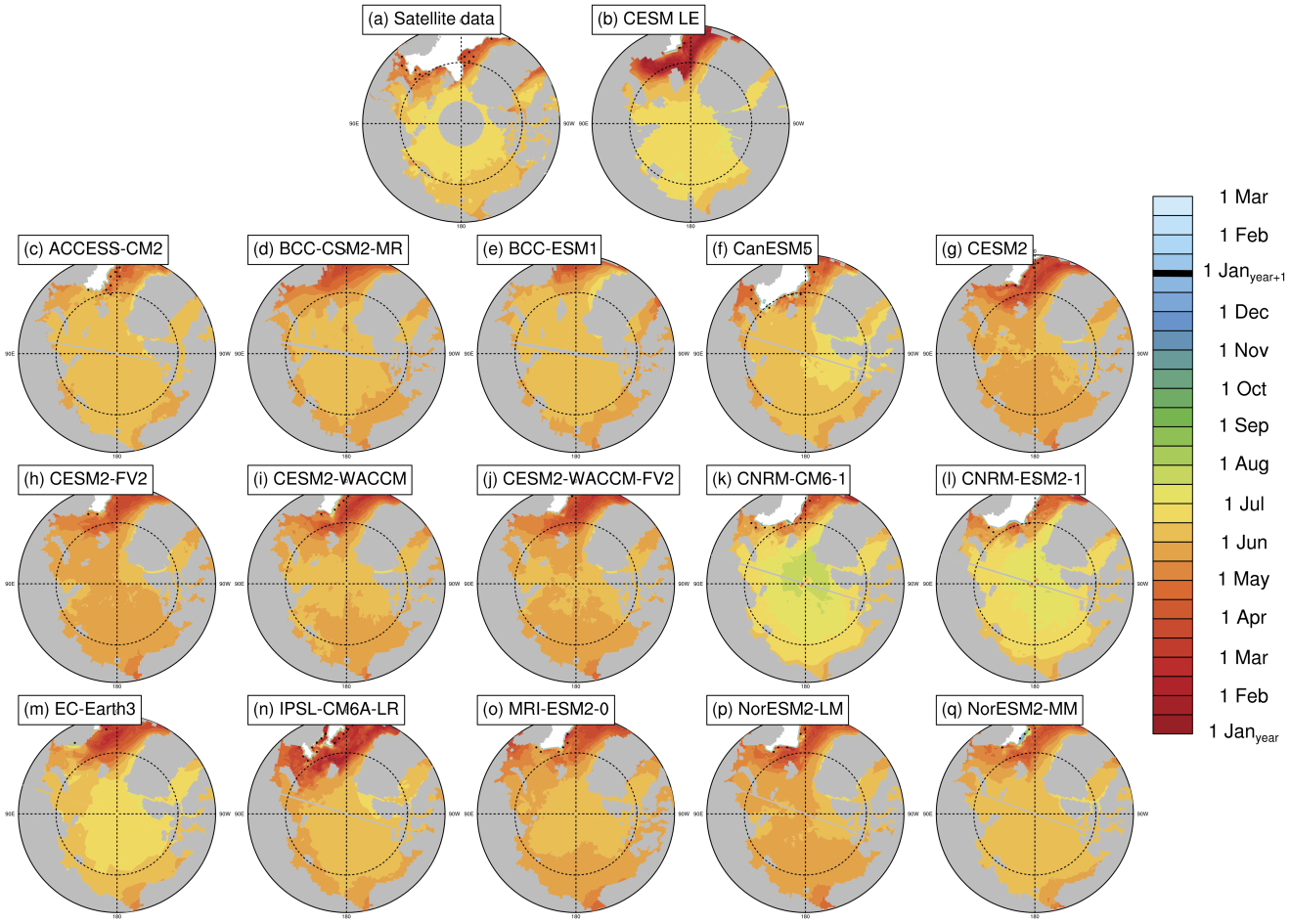

Figure 2. Melt onset dates (defined using surface temperature in the models and brightness temperatures in the satellite data) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where melt onset dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines.

~~formation~~ the start of melt on the ice or snow surface and a decline in ice area (the melt period). Once open water is present in the grid cell, the transition between 80% ice concentration and 15% ice concentration is faster, due to energy absorption from the change in the surface albedo (Perovich et al., 2002).

Models generally agree with satellite data on the timing of spring transitions (Figs. 2–4), with ~~median~~ mean melt onset dates over the satellite era occurring between May 15–June ~~17 (observed median date of May 30~~18 (observed mean date of June 6) (Table 3). Excluding the CNRM models (which show particularly late ~~median~~ mean melt onset dates and are explored further in Sect. 4.5), the model spread (May 15–June 3) shifts earlier~~toward the~~ , and the mean melt onset dates from the remaining models all occur earlier than the satellite data. The models generally capture the spatial variation in the opening and break-up

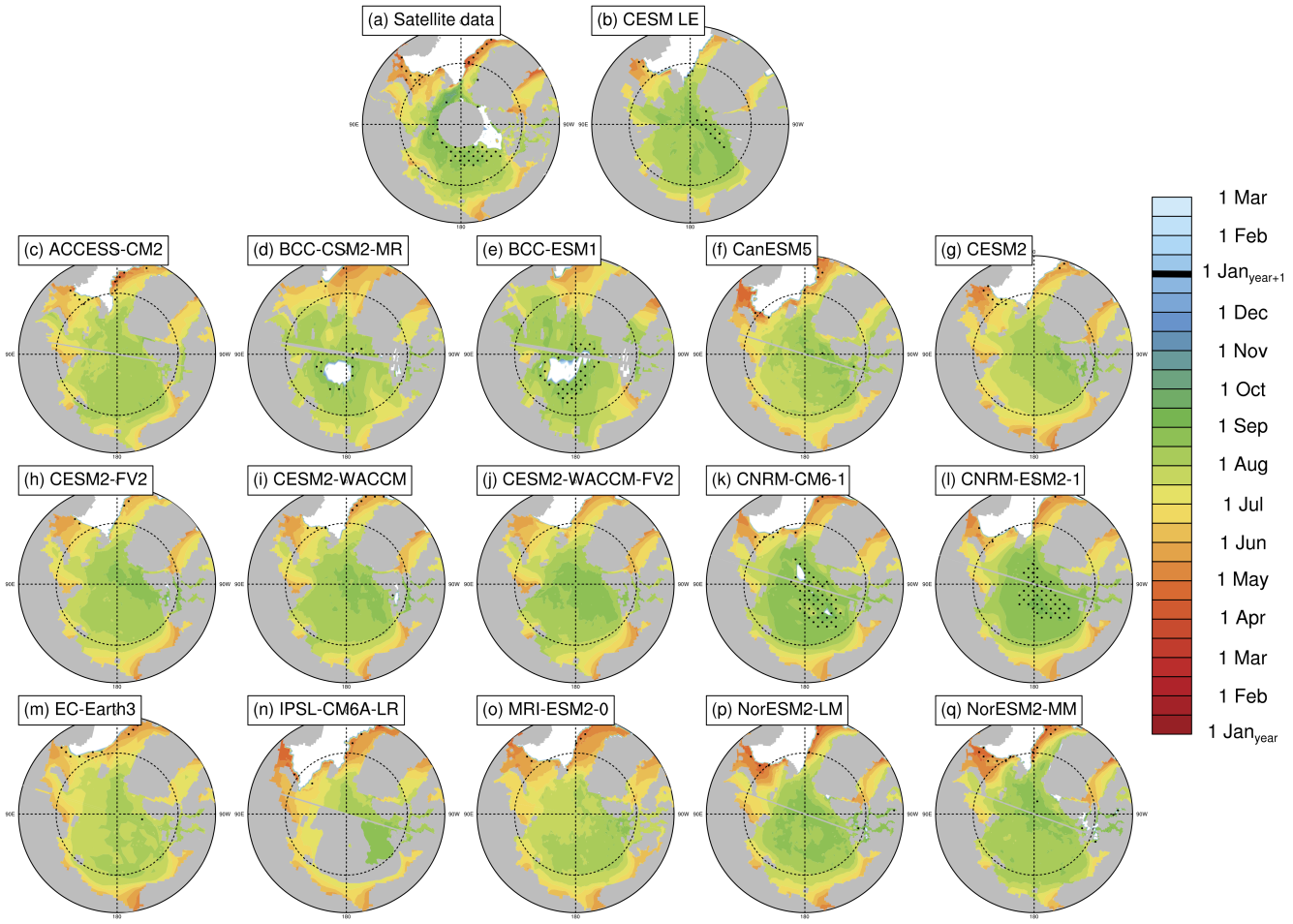

Figure 3. Opening dates (80% ice concentration threshold) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where opening dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines.

dates across the Arctic, showing a spatial standard deviation between 23–34 days compared to 32 days in satellite data and 18–28 days compared to 27 days in satellite data, respectively (Supplementary Fig. S3). However for the melt onset, the spatial standard deviation between models shows a difference by a factor of four between the models (12–47 days), compared to 20 days in the satellite data, due to the very early melt onsets detected in the Atlantic inflow regions in some models (Fig. 2). Melt onset dates in the Atlantic inflow regions that fall between January-March demonstrate that melt onset at the surface of the snow pack can occur while the ice area is still expanding in those regions. This highlights that surface temperature based definitions, such as melt onset, capture different physical processes than sea ice concentration-based definitions, as was previously shown (Smith and Jahn, 2019).

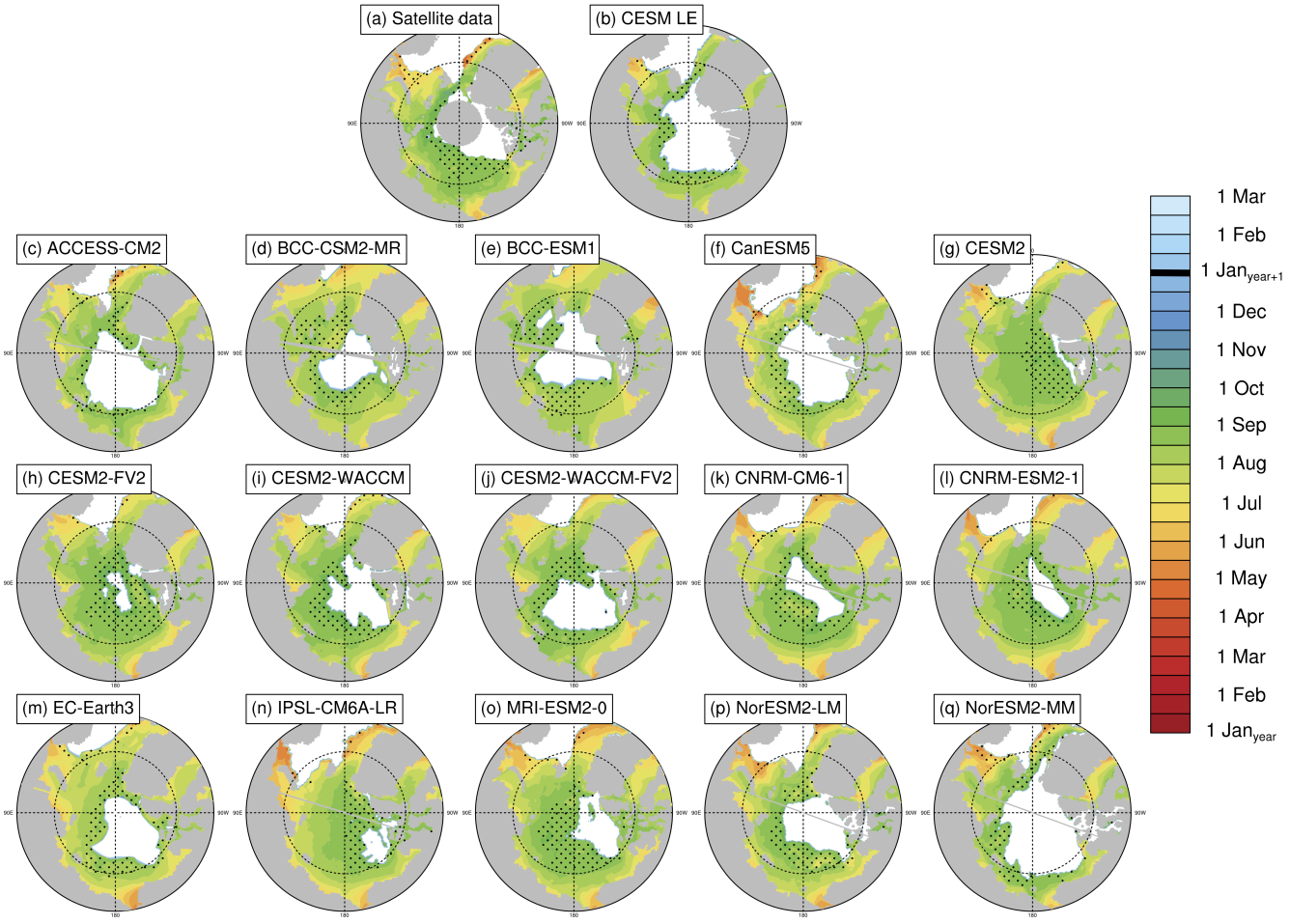

Figure 4. Break-up dates (15% ice concentration threshold) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where break-up dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines.

In agreement with observations, all models project the ~~median~~ mean length of the melt period to be longer than the seasonal loss-of-ice period ~~. The median~~ (Table 4). The mean time between melt onset and opening (the melt period) is ~~28–53~~ 32–54 days in models and ~~31~~ 39 days in the satellite data, while the ~~median~~ mean time between opening and break-up (the seasonal loss-of-ice period) is ~~13–33~~ 14–34 days in models and ~~20~~ 28 days in observations ~~.~~ (Table 4, Supplementary Figs. S4 and S5).

We find that for all spring transition metrics, the model spread exceeds estimations of internal variability, which show a maximum of ~~6~~ 8 days between ensemble members (Table 3). Of the spring sea ice transition dates (melt onset, opening and break-up), the sea ice melt onset dates show the largest spread in satellite-era ~~medians~~ means between the models (~~33~~ 34 days) (Table 3). This range is skewed late by the CNRM-ESM2-1 and CNRM-CM6-1. If the two CNRM models are excluded, the

	Melt onset	Opening (80%)	Break-up (15%)	Freeze onset	Freeze-up (15%)	Closing (80%)
ACCESS-CM2	Jun ~~2~~ 3	Jul ~~14~~ 15	Jul 31	Oct ~~3~~ 6	Oct 15	Oct ~~11~~12
BCC-CSM2-MR	May 24	Jul 15	Jul 27	Oct ~~7~~ 8	Oct 11	Oct 10
BCC-ESM1	May ~~29~~ 30	Jul 22	~~Jul 31~~ Aug 1	Oct ~~1~~ 2	Oct 10	Oct ~~7~~8
CanESM5	Jun 3	Jul ~~11~~ 12	Jul 21	Oct ~~14~~ 16	Oct 16	Oct ~~15~~16
CESM2	May 20	Jul ~~7~~ 8	Jul ~~21~~ 31	Oct 23	Oct ~~28~~ 29	Oct 29
CESM2-FV2	May ~~20~~ 22	Jul ~~12~~ 14	~~Jul 25~~ Aug 3	Oct ~~16~~ 18	Oct ~~20~~ 23	Oct ~~16~~23
CESM2-WACCM	May ~~22~~ 23	Jul ~~15~~ 16	Aug 3	Oct ~~16~~ 17	Oct 25	Oct 22
CESM2-WACCM-FV2	May ~~29~~ 21	Jul ~~19~~ 14	Jul ~~28~~ 31	Oct ~~7~~ 17	Oct ~~13~~ 26	Oct ~~6~~22
CNRM-ESM2-1	Jun ~~13~~ 14	Jul ~~18~~ 19	Jul ~~27~~ 28	Oct ~~25~~ 28	Oct ~~29~~ 30	Nov 5
CNRM-CM6-1	Jun ~~17~~ 18	Jul ~~18~~ 19	Jul ~~28~~ 29	Oct ~~18~~ 21	Oct ~~24~~ 25	Oct ~~29~~30
EC-Earth3	Jun ~~1~~ 2	Jul ~~9~~ 10	Jul 21	Oct 10	Oct 8	Oct 4
IPSL-CM6A-LR	May 15	Jul ~~5~~ 6	Jul 22	Nov ~~2~~ 3	Oct ~~23~~ 24	Oct 25
MRI-ESM2-0	May 22	Jul 6	Jul 28	Oct ~~23~~ 25	Oct 24	Oct ~~25~~26
NorESM2-LM	May 21	Jul ~~14~~ 12	~~Aug 3~~ Jul 25	Oct ~~15~~ 17	Oct ~~23~~ 21	Oct ~~22~~16
NorESM2-MM	May ~~20~~ 29	Jul ~~14~~ 19	Jul ~~31~~ 29	Oct ~~16~~ 8	Oct ~~25~~ 13	Oct ~~22~~6
CESM LE	May ~~21~~ 29	Jul ~~15~~ 22	~~Jul 28~~ Aug 5	~~Sep 28~~ Oct 7	Oct ~~14~~ 21	Oct ~~5~~13
Satellite data	~~May 30~~ Jun 6	Jul ~~8~~ 16	~~Jul 28~~ Aug 4	~~Sep 27~~ Oct 4	Oct ~~7~~ 15	Oct ~~5~~13
All-model spread	~~33~~ 34	~~17~~ 16	~~13~~ 15	~~35~~ 32	~~21~~ 22	32
CanESM5 spread*	5	4	4	~~6~~ 7	5	~~9~~8
IPSL-CM6A-LR spread*	~~6~~ 7	8	~~5~~ 6	~~10~~ 11	9	~~14~~13
CESM LE spread*	~~3~~ 4	4	5	~~6~~ 8	5	5

Table 3. Pan-Arctic, satellite-era (1979–2014) ~~medians~~ means of seasonal sea ice transition dates. The satellite-era ~~medians~~ means and the all-model spreads (latest minus earliest) are calculated using the first ensemble member from each model. Models labeled with * show the spread in ~~medians~~ means between the first 30 ensemble members of that model. Model spreads are given in days and all metrics are calculated between 66-84.5°N.

spread in ~~median~~ mean melt onset dates is ~~19~~ 22 days instead of 33 days, still larger than the other two spring metrics (~~17~~ 16 days for opening and ~~13~~ 15 days for break-up). As the CNRM melt onset dates are more than ~~a week~~ eight days later than the other models, this also means that differences between the CNRM models and the other models are unlikely explained by internal variability alone. The CNRM models are further discussed in Sect. 4.5. Of the spring transition dates, the internal
245 variability is highest for the melt onset dates, particularly in the marginal ice zones (Fig. 8). High variability between ensemble members in the marginal ice zones is likely related to the interannual variations in the position of the ice edge. Additionally,

modeled melt onset is defined using daily surface temperature, which exhibits greater variability than daily ice concentration (Smith and Jahn, 2019).

4.2 Fall transitions

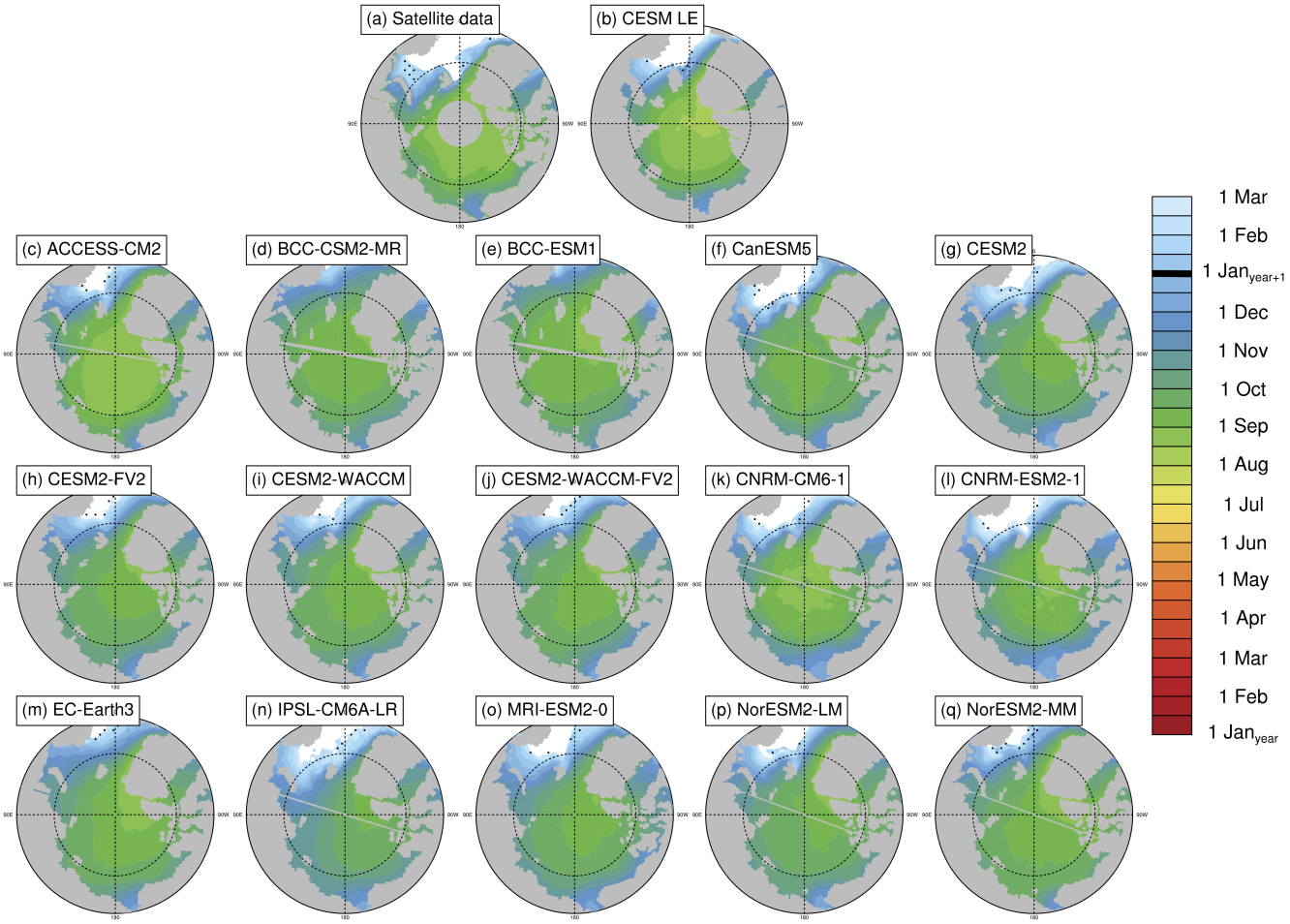

Figure 5. Freeze onset dates (defined using surface temperature in the models and brightness temperatures in the satellite data) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where freeze onset dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines.

250 Like the spring transition metrics, we find that Pan-Arctic model differences in fall transition metrics are unlikely due to internal variability alone. In the satellite data, the ~~median~~ mean freeze onset date is ~~September 27 and the median~~ October 4 and the mean freeze-up date is October ~~7.~~ 15. In the models, the satellite-era ~~medians~~ means of freeze onset fall between ~~September 28–November 2 and median~~ October 2–November 3 and mean freeze-up dates fall between October 8–October

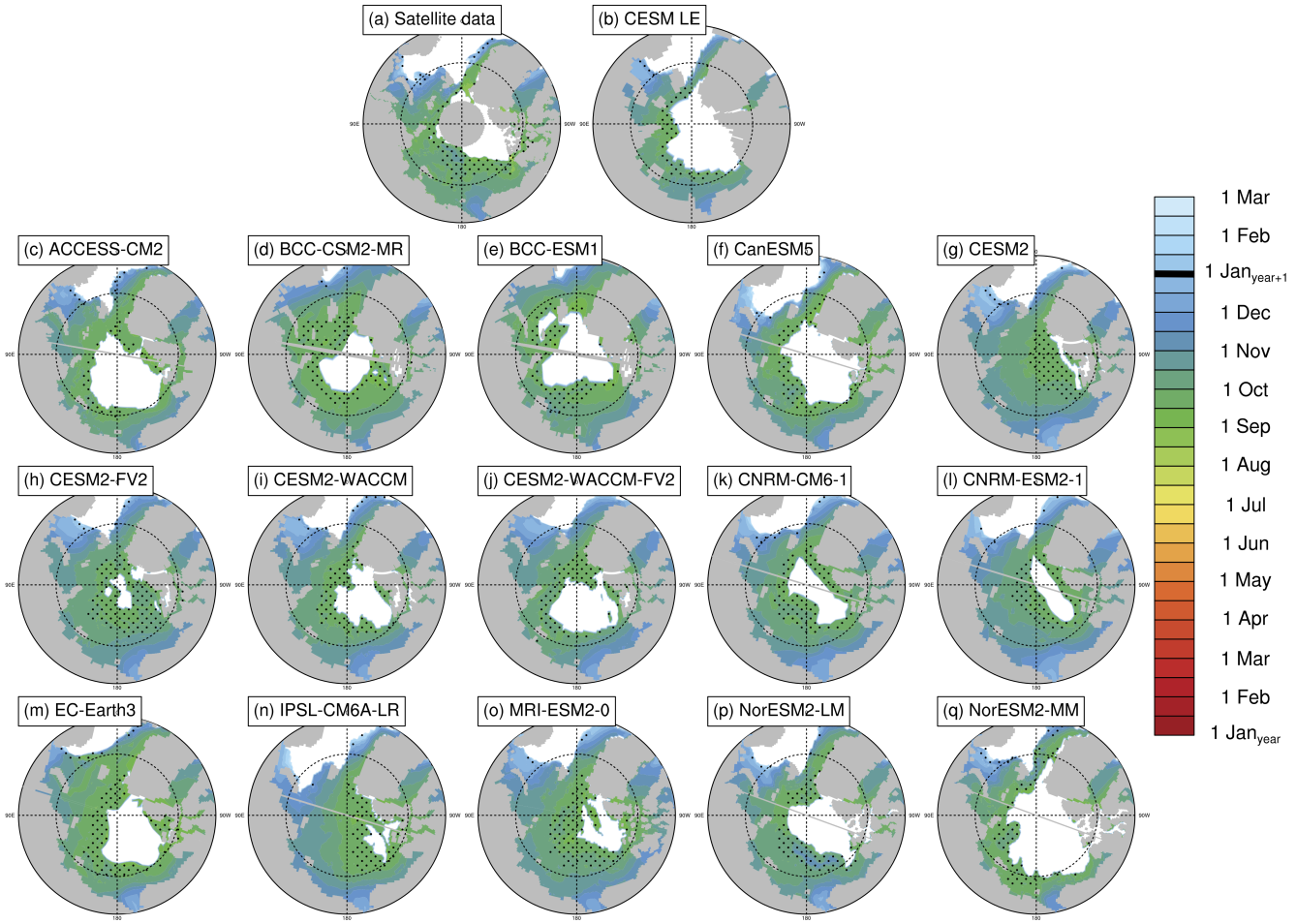

Figure 6. Freeze-up dates (15% ice concentration threshold) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where freeze-up dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines.

~~29. Models~~ 30. Multiple models therefore tend to show later freeze onset than observed (Figs. 5-7). ~~Only five of the sixteen models fall within the maximum range of internal variability (10 days) of the satellite data (Table 3)~~ The maximum range in mean freeze onset dates due to internal variability is 11 days (Table 3) and the majority of the model means (ten out of sixteen) are more than 11 days later than the satellite data, indicating that this delay of the mean freeze onset in the models is not only due to internal variability. Freeze-up is also generally delayed in models compared to satellite data, with ~~seven~~ five of the sixteen models falling ~~within~~ outside the maximum range of internal variability (9 days) of the observations. ~~The observed time between median freeze onset and freeze-up (the freeze period)is ten days, and this is twenty-nine days shorter than the time between melt onset and opening (the melt period) (Table 3)~~In the fall transition dates, the average standard deviation between

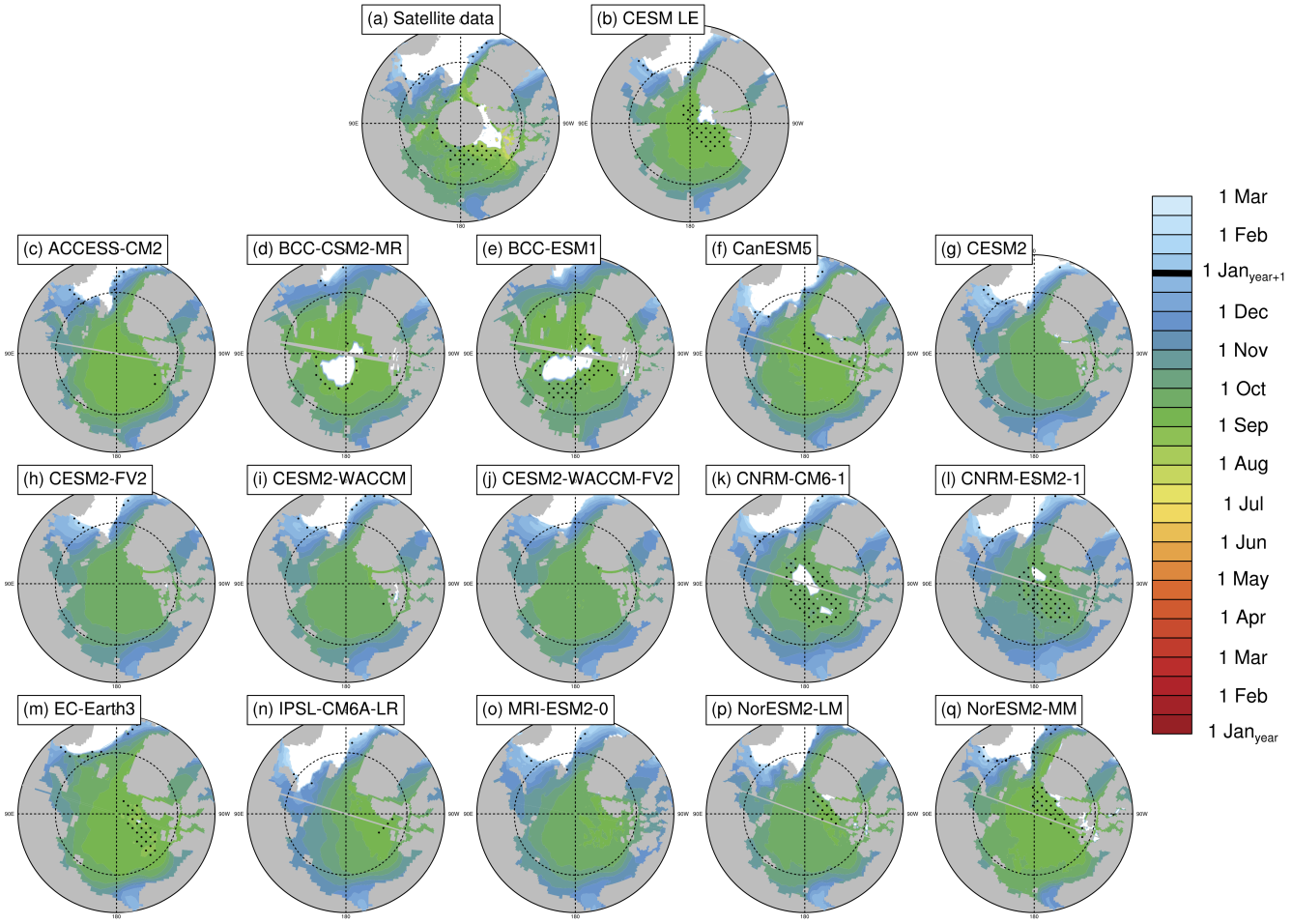

Figure 7. Closing dates (ice passes the 80% ice concentration threshold) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where closing dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines.

ensemble members is highest for marginal ice zone freeze onset dates (Fig. 8). As described for melt onset, this large internal variability is due to the changing interannual position of the ice edge and the variability of surface temperature.

The ~~sea ice closes when it passes the 80% ice concentration threshold, and the median~~ mean closing date occurs on October ~~5~~ 13 in satellite data and between October 4–November 5 in the models (Table 3). Ice freeze-up occurs before the date of closing by definition, as both are defined using ice concentration, but areas closer to the Central Arctic (that fall below 80% but not 15%) skew the ~~median~~ mean of the closing dates earlier ~~. The median~~ (in some cases earlier than the mean freeze-up date). As with the spring transition dates, models generally capture the spatial variability of the fall transition dates. Closing date standard deviations across the Arctic (which only vary by 4 days between models) do not overlap with satellite data, but freeze onset and freeze-up, which show more variation between models in terms of their standard deviations, do span the satellite data

(Table S3). The mean length of the seasonal gain-of-ice period, the time between freeze-up and closing, is ~~7~~ 15 days in satellite data and ~~4–14~~ 7–14 days in models. Thus the seasonal loss-of-ice period is almost ~~three times~~ twice as long as the seasonal gain-of-ice period~~.~~

~~Like the spring transition metrics, we find that Pan-Arctic model differences in fall transition metrics are unlikely due to internal variability alone. Of the fall sea ice transition dates (freeze onset, freeze-up and closing), freeze onset shows the largest spread in satellite-era median between models (35 days) (Table 3). Freeze-up and closing have median spreads between the models of 21 and 32 days respectively. The maximum average standard deviation between the ensemble members of CESM LE, CanESM5 and IPSL for the fall transition metrics is 14 days. Therefore the model spreads of all fall transition metrics exceed estimations of internal variability. In the fall transition dates, the average standard deviation between ensemble members is highest for marginal ice zone freeze onset dates (Fig. 8). As described for melt onset, this large internal variability is due to the changing interannual position of the ice edge and the variability of surface temperature.~~

~~The average standard deviation between the first 30 ensemble members over 1979–2014 for (a) melt onset (b) opening (c) break-up (d) freeze onset (e) freeze-up (f) closing. CanESM5 is displayed in the first row, IPSL is displayed in the second row and CESM LE is displayed in the third row. The standard deviation is calculated at each grid cell for each year, and then the average of all years is plotted for each grid cell. The same figure using all available ensemble members of each model is displayed in Supplementary Fig. S2.~~

~~The fall~~ , in both satellite data and models (Table 4, Supplementary Figs. S5 and S7).

The observed time between mean freeze onset and freeze-up (the freeze period) is nine days, with the mean freeze-up occurring before the mean freeze onset in the satellite data, leading to a negative freeze period in the mean and across most of the Arctic (Table 4, Supplementary Figs. S6). This means that the transition dates do not always occur in the expected order at each grid cell, ~~with freeze-up occurring at the same time or earlier than freeze onset.~~ and out-of-order dates occur much more frequently for the freeze period than the melt period (Supplementary Figs. S4 and S6). In satellite data, simultaneous freeze onset and freeze-up dates may in part be explained by the satellite retrieval algorithms: the PMW retrieval algorithm for freeze onset uses an 80% ice concentration metric to derive freeze onset at locations where the date can not be reliably derived using the weighted brightness temperature scheme (Markus et al., 2009)~~, but the~~ . This would skew the freeze onset dates later and make them more similar to the closing dates. Hence, the use of ice concentration by both the freeze onset and freeze-up retrieval algorithms may contribute to cases where the dates are not sequential. A detailed assessment of this is not possible, however, as the data does not contain information on how often this back-up method is employed. ~~Hence, the use of ice concentration by both the~~

In models, the definitions of seasonal sea ice metrics aim to capture thermodynamic changes in the sea ice, but the similar and sometimes out-of-order dates for freeze onset and freeze-up ~~retrieval algorithms may contribute to cases where the dates are equal.~~ highlight that dynamic sea ice changes influence the ice concentration-based transition metrics as well. While a particular grid cell may not register a persistent change in surface temperature below the threshold for freeze onset (-1.8 °C), it is possible that the ice concentration of the grid cell surpasses 15% due to dynamic transport into the grid cell, triggering the

detection of freeze-up. This occurs in several models such that freeze onset occurs later than freeze-up in some parts of Arctic, leading to negative freeze periods, as found for the satellite data (Supplementary Figure S6).

	Melt period	Seasonal loss-of-ice period	Freeze period	Seasonal gain-of-ice period
ACCESS-CM2	43	30	11	11
BCC-CSM2-MR	52	22	-1	9
BCC-ESM1	54	20	1	10
CanESM5	40	26	-1	14
CESM2	45	32	11	7
CESM2-FV2	52	33	10	7
CESM2-WACCM	52	33	9	8
CESM2-WACCM-FV2	53	31	11	8
CNRM-ESM2-1	37	14	4	12
CNRM-CM6-1	32	17	6	12
EC-Earth3	33	25	-4	10
IPSL-CM6A-LR	42	25	-8	9
MRI-ESM2-0	43	34	-1	14
NorESM2-LM	52	31	4	8
NorESM2-MM	52	33	4	9
CESM LE	39	30	-7	11
Satellite data	39	28	-9	15
All-model spread	22	20	19	7
CanESM5 spread*	4	3	4	1
IPSL-CM6A-LR spread*	5	2	8	1
CESM LE spread*	4	3	8	2

Table 4. Lengths of pan-Arctic, satellite-era (1979–2014) means of intra-seasonal transition periods in days. The satellite-era means and the all-model spreads (latest minus earliest) are calculated using the first ensemble member from each model. Negative values indicate where the freeze-up date falls earlier than the freeze onset on average. Models labeled with * show the spread in means between the first 30 ensemble members of that model. Model spreads are given in days and all metrics are calculated between 66-84.5°N.

~~In models, the definitions of seasonal sea ice metrics aim to capture thermodynamic changes in the sea ice, but the similar dates for freeze onset and freeze-up highlight that dynamic sea ice changes influence the ice concentration-based seasonal sea ice transition metrics as well. While a particular grid cell may not register a persistent change in surface temperature below the threshold for freeze onset (-1.8 C), it is possible that the ice concentration of the grid cell surpasses 15% due to dynamic transport into the grid cell, triggering the detection of freeze-up. In multiple models, freeze onset occurs later than freeze-up in~~

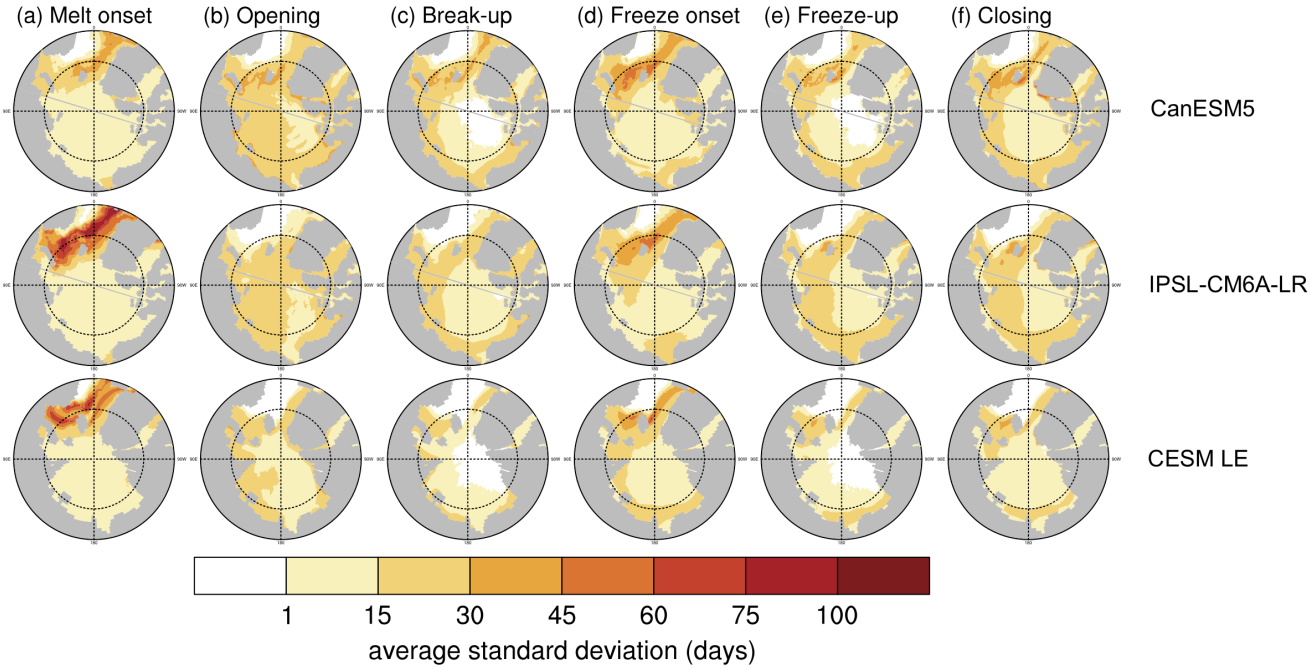

Figure 8. The average standard deviation between the first 30 ensemble members over 1979–2014 for (a) melt onset (b) opening (c) break-up (d) freeze onset (e) freeze-up (f) closing. CanESM5 is displayed in the first row, IPSL is displayed in the second row and CESM LE is displayed in the third row. The standard deviation is calculated at each grid cell for each year, and then the average of all years is plotted for each grid cell. The same figure using all available ensemble members of each model is displayed in Supplementary Fig. S2.

~~some parts of the Central Arctic. All models show freeze onset later than freeze-up occurring in the marginal ice zones, which are particularly susceptible to dynamic ice changes due to generally lower ice concentrations.~~

4.3 Inter-seasonal transition periods

Out of the three inter-seasonal periods of transition (the melt season, the open water period and the outer ice-free period), the outer ice-free period is the only one that is consistent with satellite data. The outer ice-free period (80% ice concentration thresholds) has an observed ~~median length of 81~~ mean length of 88 days and model ~~medians falling between 72–95~~ means falling between 73–95 days (Table 5 and Supplementary Fig. S8).

In contrast, the melt season length and the open water period are too long in models compared to observations. Generally, the greatest contribution to the differences between the observed and modeled open water period is from later than observed freeze-up dates. The open water period has a ~~median of 64~~ mean of 71 days in the satellite data and ~~medians ranging between 73–113~~ means ranging between 78–113 days in the models (a spread of ~~40~~ 35 days) (Table 5 ~~). Likewise, modeled~~ and Supplementary Fig. S9). Modeled melt seasons that are too long compared to observations are also largely driven by ~~later~~ their fall transition metric (freeze onset dates) occurring late. The observed ~~median~~ mean melt season length is ~~110~~ 117 days

	Melt season	Open water period (15%)	Outer ice-free period (80%)
ACCESS-CM2	121	~~88~~ 89	~~77~~78
BCC-CSM2-MR	136	87	~~77~~78
BCC-ESM1	~~124~~ 125	78	~~72~~73
CanESM5	~~129~~ 130	95	88
CESM2	154	113	~~90~~91
CESM2-FV2	~~147~~ 148	~~96~~ 101	~~88~~83
CESM2-WACCM	~~146~~ 147	~~98~~ 99	84
CESM2-WACCM-FV2	~~130~~ 148	~~77~~ 100	~~76~~88
CNRM-ESM2-1	132	110	95
CNRM-CM6-1	~~120~~ 121	103	89
EC-Earth3	127	~~85~~ 86	80
IPSL-CM6A-LR	165	111	95
MRI-ESM2-0	~~153~~ 154	111	90
NorESM2-LM	~~148~~ 147	~~100~~ 96	~~82~~89
NorESM2-MM	~~147~~ 130	~~100~~ 78	~~88~~77
CESM LE	~~115~~ 125	~~73~~ 80	~~75~~82
Satellite data	~~110~~ 117	~~64~~ 71	~~81~~88
All-model spread	~~50~~ 44	~~40~~ 35	~~23~~22
CanESM5 spread*	~~9~~ 10	7	12
IPSL-CM6A-LR spread*	17	11	21
CESM LE spread*	~~9~~ 12	9	9

Table 5. Lengths of pan-Arctic, satellite-era (1979–2014) ~~medians~~ means of inter-seasonal transition periods in days. The satellite-era ~~medians~~ means and the all-model spreads (latest minus earliest) are calculated using the first ensemble member from each model. Models labeled with * show the spread in ~~medians~~ means between the first 30 ensemble members of that model. Model spreads are given in days and all metrics are calculated between 66-84.5°N.

and the model ~~medians range between 115-165~~ means range between 121-165 days (Table 5 and Supplementary Fig. S10). Therefore the melt season length exhibits the largest model spread of all the inter-seasonal periods (~~50~~ 44 days). This is due to larger model ranges in both melt onset and freeze onset than the other transition dates, and the contribution of each date to the model spread in melt season length is approximately equal (Table 5). The melt season length model range is also skewed high by the IPSL model, which has a ~~median~~ mean melt season length of 165 days in its first ensemble member. This is 11 days longer than the next longest model ~~median~~mean, and the choice of ensemble member likely plays a role—the IPSL model has a particularly large range of internal variability in the ~~median~~ mean melt season length (17 days compared to ~~9~~ 10 and 12

days in the other two model sets) (Table 5). While the mean melt season and open water periods are long compared to satellite data, their modeled spatial standard deviations agree with satellite data, while those of the outer-ice free period are all slightly smaller than observed (Table S4).

	Melt onset	Opening (80%)	Break-up (15%)	Freeze onset	Freeze-up (15%)	Closing (80%)
ACCESS-CM2	~~0.66[a]~~ **0.66**	~~0.78[a]~~ **0.78**	0.22	~~-0.84[a]~~ **-0.84**	~~-0.77[a]~~ **-0.77**	~~-0.73[a]~~ **-0.73**
BCC-CSM2-MR	~~0.47[a]~~ **0.47**	~~0.64[a]~~ **0.64**	~~0.39[a]~~ **0.39**	~~-0.80[a]~~ **-0.80**	~~-0.60[a]~~ **-0.60**	~~-0.74[a]~~ **-0.74**
BCC-ESM1	~~0.53[a]~~ **0.53**	-0.14	-0.05	~~-0.67[a]~~ **-0.67**	~~-0.53[a]~~ **-0.53**	~~-0.45[a]~~ **-0.45**
CESM2	~~0.37[a]~~ **0.37**	~~0.87[a]~~ **0.87**	~~0.43[a]~~ **0.43**	~~-0.87[a]~~ **-0.87**	~~-0.78[a]~~ **-0.78**	~~-0.87[a]~~ **-0.87**
CESM2-FV2	~~0.70[a]~~ **0.70**	~~0.89[a]~~ **0.89**	0.21	~~-0.90[a]~~ **-0.90**	~~-0.75[a]~~ **-0.75**	~~-0.82[a]~~ **-0.82**
CESM2-WACCM	~~0.62[a]~~ **0.62**	~~0.85[a]~~ **0.85**	0.22	~~-0.86[a]~~ **-0.86**	~~-0.68[a]~~ **-0.68**	~~-0.79[a]~~ **-0.79**
CESM2-WACCM-FV2	~~0.61[a]~~ **0.61**	~~0.81[a]~~ **0.81**	~~0.47[a]~~ **0.47**	~~-0.84[a]~~ **-0.84**	~~-0.73[a]~~ **-0.73**	~~-0.78[a]~~ **-0.78**
CNRM-ESM2-1	0.08	-0.32	-0.25	0.24	0.14	0.14
CNRM-CM6-1	0.13	-0.19	-0.15	-0.07	-0.04	-0.12
EC-Earth3	~~0.85[a]~~ **0.85**	~~0.65[a]~~ **0.65**	~~0.53[a]~~ **0.53**	~~-0.93[a]~~ **-0.93**	~~-0.83[a]~~ **-0.83**	~~-0.85[a]~~ **-0.85**
MRI-ESM2-0	~~0.49[a]~~ **0.49**	~~0.66[a]~~ **0.66**	-0.05	~~-0.91[a]~~ **-0.91**	~~-0.82[a]~~ **-0.82**	~~-0.84[a]~~ **-0.84**
NorESM2-LM	~~0.55[a]~~ **0.55**	~~0.69[a]~~ **0.69**	0.11	~~-0.73[a]~~ **-0.73**	~~-0.51[a]~~ **-0.51**	~~-0.64[a]~~ **-0.64**
NorESM2-MM	~~0.56[a]~~ **0.56**	0.18	-0.27	~~-0.74[a]~~ **-0.74**	~~-0.61[a]~~ **-0.61**	~~-0.49[a]~~ **-0.49**
CanESM5	~~0.73[a]~~ **0.74**	~~0.65[a]~~ **0.66**	~~0.11[a]~~ **0.11**	~~-0.83[a]~~ **-0.84**	~~-0.68[a]~~ **-0.68**	~~-0.71[a]~~ **-0.71**
IPSL-CM6A-LR	~~0.57[a]~~ **0.58**	~~0.70[a]~~ **0.70**	~~0.12[a]~~ **0.13**	~~-0.88[a]~~ **-0.88**	~~-0.80[a]~~ **-0.80**	~~-0.84[a]~~ **-0.84**
CESM LE	~~0.54[a]~~ **0.54**	~~0.41[a]~~ **0.41**	~~0.06[a]~~ **0.06**	~~-0.87[a]~~ **-0.87**	~~-0.43[a]~~ **-0.43**	~~-0.56[a]~~ **-0.56**
Satellite data	~~0.81[a]~~ **0.83**	~~0.72[a]~~ **0.72**	~~0.65[a]~~ **0.64**	~~-0.93[a]~~ **-0.93**	~~-0.64[a]~~ **-0.64**	~~-0.81[a]~~ **-0.82**

Table 6. Correlation coefficients (R-values) between seasonal sea ice transition dates and mean summer (June–September) sea ice area of the same year from 1979–2014. Values ~~with [a]~~ in bold are statistically significant at the 95% confidence level. Correlation coefficients and p-values for models in the first thirteen rows are determined using one ensemble member, for CanESM5 using all 35 ensemble members, for IPSL using all 30 ensemble members and CESM LE using all 40 ensemble members. All values are calculated between 66-84.5°N.

4.4 Seasonal transitions affect sea ice area and thickness year-round

Model representations of seasonal sea ice transitions are expected to impact sea ice area and thickness because seasonal transitions are strongly linked to the ice-albedo feedback (Perovich et al., 2008; Timmermans, 2015; Kashiwase et al., 2017; Perovich, 2018; Lebrun et al., 2019). Ice loss earlier in the spring has been related to later ice gain in the fall (Stroeve et al., 2014, 2016; Lebrun et al., 2019), and a weaker relationship has been described between later ice gain and earlier spring loss during the following year (Lebrun et al., 2019). Both processes favor greater areas of open ocean for longer periods each year, but little has been done to evaluate which transition metrics are most appropriate for describing pan-Arctic sea ice relationships.

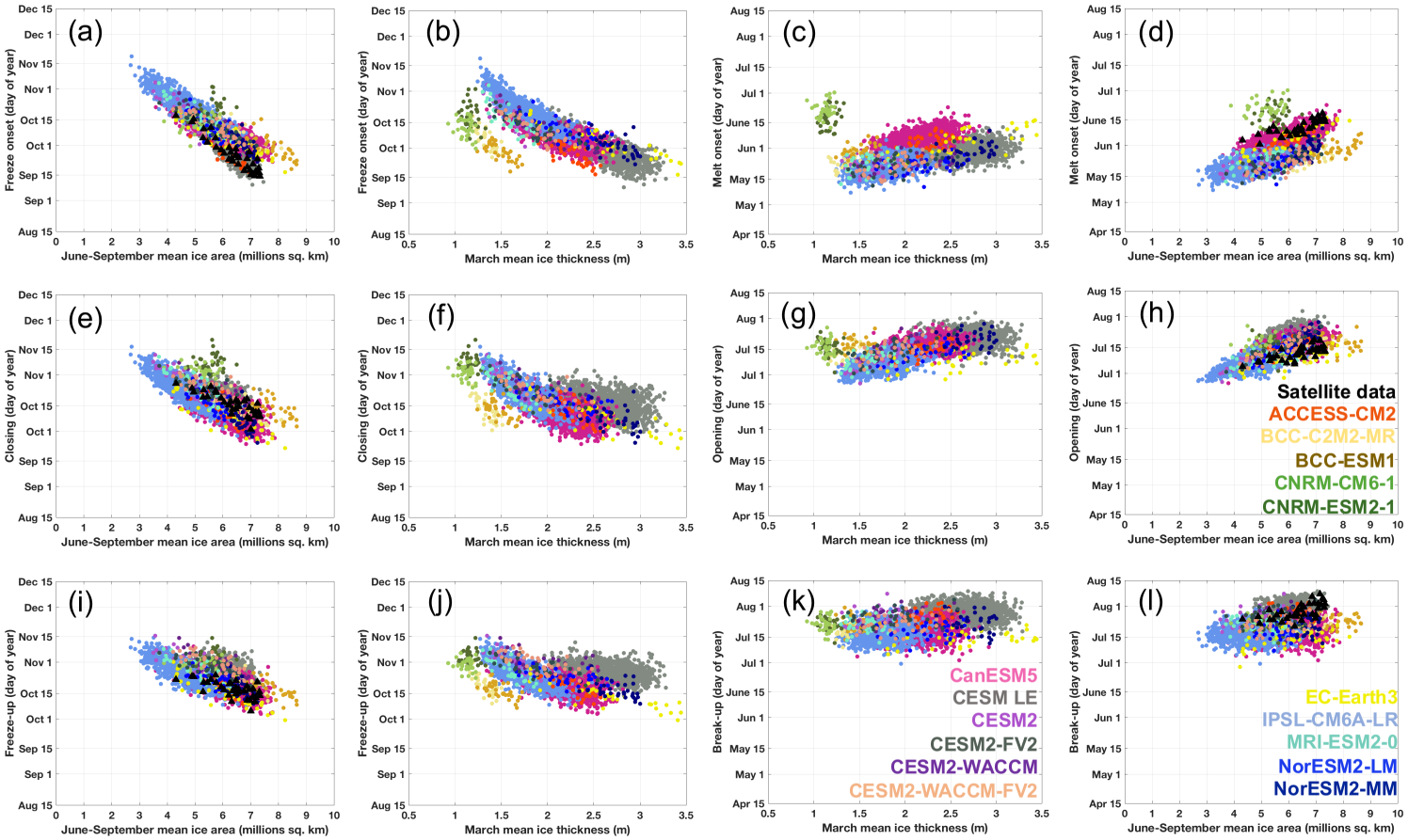

Figure 9. Scatter plots of mean seasonal sea ice transition metrics versus other ice characteristics for CMIP6 models (colors), CESM LE (gray) and satellite data (black). Each scatter point represents one year in one ensemble member from 1979–2014. Panels (a-d) show relationships with mean melt and freeze onset dates, panels (e-h) show relationships with mean opening and closing dates and panels (i-l) show relationships with mean break-up and freeze-up dates. Metrics are scattered against mean summer (June–September) ice area in the first and fourth columns and March mean ice thickness in the second and third columns. All metrics are scattered against ice characteristics from the same year, except those in the second column, in which fall transition metrics are scattered against the next year's mean March ice thickness. All available ensemble members are used for CESM LE, CanESM5 and IPSL. All metrics are calculated between 66-84.5°N. All models are represented in each panel (a)-(l), but the labels are distributed across panels (h), (k) and (l).

Here we demonstrate year-round relationships using seasonal transition dates, March mean ice thickness and summer (June-September) mean ice area. We show that pan-Arctic relationships between seasonal transitions and other ice characteristics are most discernible using seasonal transition metrics with extensive spatial coverage (Fig. 9). Summer mean ice area is evaluated instead of the ice area of a single month in order to better represent the integrated surface energy absorption as ice area declines. Ice area and seasonal ice transition dates are practical for assessing sea ice in a pan-Arctic sense, as they are reliably available for both models and observations. Discussion of the sea ice thickness here is limited to model projections, since observations of Arctic sea ice thickness are temporally limited and contain large uncertainties (Bunzel et al., 2018).

We find that in satellite data, mean summer ice area (June–September) and the ~~median~~ mean timing of freeze onset are strongly anti-correlated (R=-0.93) (Table 6 and Fig. 9). Lower summer ice area corresponds to a lower surface albedo, allowing for greater shortwave absorption by the surface ocean and increasing ocean heat content (Timmermans, 2015), delaying the freeze onset (Stroeve et al., 2014). Slightly weaker relationships exist between mean observed summer sea ice area and freeze-up (R=-0.64) and closing (R=~~-0.81~~ -0.82). In models, the greatest agreement on the correlation between mean summer ice area and fall transition metrics is seen using the freeze onset dates, where all of the correlation coefficients that are statistically significant at the 95% level (~~13~~ 14 out of 16 models) are equal to or more negative than ~~-0.73~~ -0.67 (Table 6). Models tend to show later freeze onset than observed, as discussed in Sect. 4.2, and despite this offset the observed relationship between summer ice area and freeze onset is captured well by the models (Fig. 9). Summer ice area is generally larger in satellite data than in the models, but falls within the model spread. Relationships between fall transition dates and mean summer ice thickness are similar but slightly weaker than found with ice area (Supplementary Table S5).

In the models, the timing of fall transition dates are strongly correlated with the March mean ice thickness (Table 7 and Fig. 9), but do not affect the March ice area of the following year (Supplementary Table ~~S3~~S6). The differences in correlation coefficients indicate that increased heat absorption and delayed freeze onset reduce the March thickness of the ice but have a much smaller impact on the ice area. This supports past work on the Canada Basin, showing that anomalous solar heat input (Perovich et al., 2008) reduced ice thickness over the winter of 2007-2008 by 25% (Timmermans, 2015). The strongest correlations between March ice thickness and the previous year's fall transition metrics are found between freeze onset and March ice thickness, with statistically significant correlation coefficients ranging between -0.54 and -0.92 in the models (Table 7). Because of sea ice thickness uncertainties discussed earlier (Bunzel et al., 2018), we are unable to confidently evaluate whether model biases in freeze onset impact the simulated relationship between freeze onset and March mean ice thickness compared to observations. With respect to the other fall transition metrics, we find that statistically significant correlations between March ice thickness and freeze-up/closing (which are both based on ice concentration) are less consistent between models, and generally stronger for the closing dates rather than freeze-up dates (Table 7). Additionally, other relationships involving freeze-up and spring sea ice of the following year (such as the relationship between the timing of freeze-up and the next year's break-up) have been shown to be dampened by the tendency of thin ice to grow faster than thicker ice (Bitz and Roe, 2004; Lebrun et al., 2019). The growth rate of thin ice, in addition to the spatial coverage of the freeze-up dates, may be limiting the impact that a late freeze-up date has in reducing the following year's March mean ice thickness.

	Melt onset	Opening (80%)	Break-up (15%)	Freeze onset	Freeze-up (15%)	Closing (80%)
ACCESS-CM2	0.25	~~0.64[a]~~ **0.64**	0.3	~~-0.76[a]~~ **-0.76**	~~-0.79[a]~~ **-0.79**	~~-0.67[a]~~ **-0.67**
BCC-CSM2-MR	~~0.43[a]~~ **0.43**	~~0.55[a]~~ **0.55**	~~0.35[a]~~ **0.35**	~~-0.82[a]~~ **-0.82**	~~-0.65[a]~~ **-0.65**	~~-0.77[a]~~ **-0.77**
BCC-ESM1	~~0.49[a]~~ **0.49**	-0.12	0.02	~~-0.68[a]~~ **-0.68**	~~-0.60[a]~~ **-0.60**	~~-0.52[a]~~ **-0.52**
CESM2	0.30	~~0.72[a]~~ **0.72**	0.29	~~-0.88[a]~~ **-0.88**	~~-0.84[a]~~ **-0.84**	~~-0.91[a]~~ **-0.91**
CESM2-FV2	~~0.60[a]~~ **0.60**	~~0.65[a]~~ **0.65**	0.13	~~-0.88[a]~~ **-0.88**	~~-0.73[a]~~ **-0.73**	~~-0.76[a]~~ **-0.76**
CESM2-WACCM	~~0.48[a]~~ **0.48**	~~0.62[a]~~ **0.62**	-0.06	~~-0.79[a]~~ **-0.79**	~~-0.54[a]~~ **-0.54**	~~-0.72[a]~~ **-0.72**
CESM2-WACCM-FV2	~~0.48[a]~~ **0.48**	~~0.65[a]~~ **0.65**	~~0.38[a]~~ **0.38**	~~-0.81[a]~~ **-0.81**	~~-0.70[a]~~ **-0.70**	~~-0.74[a]~~ **-0.74**
CNRM-ESM2-1	0.32	-0.26	-0.17	-0.17	-0.04	-0.07
CNRM-CM6-1	0.09	-0.10	-0.12	-0.23	-0.17	-0.24
EC-Earth3	~~0.75[a]~~ **0.75**	~~0.57[a]~~ **0.57**	~~0.49[a]~~ **0.49**	~~-0.92[a]~~ **-0.92**	~~-0.85[a]~~ **-0.85**	~~-0.79[a]~~ **-0.79**
MRI-ESM2-0	~~0.43[a]~~ **0.43**	~~0.42[a]~~ **0.42**	-0.13	~~-0.84[a]~~ **-0.84**	~~-0.83[a]~~ **-0.83**	~~-0.82[a]~~ **-0.82**
NorESM2-LM	~~0.42[a]~~ **0.42**	~~0.60[a]~~ **0.60**	0.07	~~-0.74[a]~~ **-0.74**	~~-0.53[a]~~ **-0.53**	~~-0.63[a]~~ **-0.63**
NorESM2-MM	~~0.38[a]~~ **0.38**	0.01	~~-0.37[a]~~ **-0.37**	~~-0.54[a]~~ **-0.54**	~~-0.41[a]~~ **-0.41**	~~-0.37[a]~~ **-0.37**
CanESM5	~~0.64[a]~~ **0.65**	~~0.51[a]~~ **0.52**	0.01	~~-0.73[a]~~ **-0.73**	~~-0.64[a]~~ **-0.64**	~~-0.64[a]~~ **-0.64**
IPSL-CM6A-LR	~~0.39[a]~~ **0.41**	~~0.54[a]~~ **0.54**	~~0.09[a]~~ **0.1**	~~-0.88[a]~~ **-0.89**	~~-0.77[a]~~ **-0.78**	~~-0.80[a]~~ **-0.80**
CESM LE	~~0.26[a]~~ **0.26**	~~0.14[a]~~ **0.13**	~~-0.11[a]~~ **-0.12**	~~-0.79[a]~~ **-0.79**	~~-0.39[a]~~ **-0.39**	~~-0.45[a]~~ **-0.45**

Table 7. Correlation coefficients (R-values) between seasonal sea ice transition dates and March sea ice thickness from 1979-2014. Spring transition dates (melt onset, opening and break-up) are correlated with March mean ice thickness from the same year, while fall transition dates (freeze onset, freeze-up and closing) are correlated with March mean ice thickness from the following year. Values ~~with [a]~~ in bold are statistically significant at the 95% confidence level. Correlation coefficients and p-values for models in the first thirteen rows are determined using one ensemble member, for CanESM5 using all 35 ensemble members, for IPSL using all 30 ensemble members and CESM LE using all 40 ensemble members. All values are calculated between 66-84.5°N.

Modeled melt onset and opening dates both demonstrate weak to moderate relationships with the mean March ice thickness of the same year (Table 7 and Fig. 9). Thinner March sea ice generally corresponds with earlier ~~median~~ mean melt onset and opening dates. For March ice thickness and melt onset, statistically significant correlations range from 0.26-0.75, with the CESM LE representing the weakest relationship in that range (Table 7). For March ice thickness and opening dates, statistically significant correlations range from ~~0.14~~0.13-0.65 (Table 7). One might expect that thinner ice would correspond to earlier break-up dates, because thinner ice is easier to melt out or split apart. However, models do not agree on the sign or statistical significance of any relationship between break-up (which are defined using ice concentration, like opening dates) and March mean ice thickness. This lack of relationship is a strong indication that the spatial coverage of break-up dates is not sufficient for describing pan-Arctic sea ice feedbacks. ~~Relationships between~~ Increases in ice thickness after March may dampen the relationship between thin March ice and an earlier break-up date, since some models show faster ice growth from March to April in areas of thin March ice rather than thick March ice (supporting past work on ice growth rates (Bitz and Roe, 2004)).

However, this pattern is not seen in all models and thus cannot fully account for the weakness of the relationships between March ice thickness and break-up. In addition, relationships between spring transition dates and March ice area are weaker than those between spring transition dates and March ice thickness (Supplementary Table S6).

Melt onset and opening are related to mean summer ice area in both satellite data and models (Table 6 and Fig. 9) (excluding the CNRM models, which are discussed in Sect. 4.5). Earlier melt is correlated with lower mean summer ice area with a correlation coefficient of ~~0.81~~ 0.83 in satellite data and a range of 0.37-0.85 in statistically significant model correlations (Table 7). Earlier opening is slightly less correlated with lower mean summer ice area with a correlation coefficient of 0.72 in satellite data, and the models range between 0.41-0.89 in statistically significant correlations (Table 7). Both earlier melt onset and opening dates decrease the surface albedo—the former ~~though~~ through the formation of melt ponds and the latter through the presence of more open ocean. This once again facilitates greater surface absorption, which has been shown to increase the ocean heat content and decrease the summer sea ice cover (Stroeve et al., 2014). Relationships also exist between melt onset/opening and summer sea ice thickness (Supplementary Table S5), but since the summer sea ice is already quite thin, greater ocean heat content is more likely to affect the ice area than it is in March, when ice is much thicker overall. Models do not agree on the sign or magnitude of the correlation between break-up and summer ice area, again indicating that the spatial coverage of break-up dates is insufficient for describing pan-Arctic sea ice processes.

4.5 Seasonal transitions can compensate for unrealistic sea ice characteristics

CNRM-CM6-1 and CNRM-ESM2-1 demonstrate that biases in seasonal sea ice transitions can unrealistically compensate for other sea ice biases. As mentioned in Sect. 4.1 and 4.4, the CNRM models show ~~median~~ mean melt onset dates occurring ~~10-14~~ 11-15 days later than the next latest model and ~~considerably later (14-18 days)~~ 8-12 days later than those found in satellite data (Table 5). The largest differences in melt onset between the CNRM models and both satellite data and the other models are found in the Central Arctic (Fig. 2k, l). While melt onset dates fall late in the CNRM models, their September ice areas are overall realistic (Voldoire et al., 2019) and fall within the spread of available models (Supplementary Fig. S1). The CNRM models are the only two models (out of sixteen) that ~~the~~ lack statistically significant correlations between later melt onset and larger summer ice area seen in most models and observations (Sect. 4.4). Furthermore, mean ice thickness in the CNRM models from 1979-2014 is ~~too low~~ much lower than in any of the other models (Supplementary Fig. S1). Thus, the models' ability to produce realistic September sea ice areas likely relies on the biased seasonal transition: late melt onset acts to retain thin ice that would otherwise be lost over the summer by shortening the length of the melt season. Following melt onset, the CNRM models have ~~median~~ mean opening and break-up dates that fall fully within the model spread, indicating that the impact of seasonal transition biases can be ~~be~~ large, even if the biases exist only in one metric.

The cause of delayed melt onset in CNRM models is not currently clear. Melt onset is the only transition metric that captures changes at the surface of the snowpack rather than a change in ice concentration. Recent work suggests that the winter snow on the sea ice is too thick in the CNRM models~~(Voldoire et al., 2019)~~, over-insulating the sea ice and preventing it from reaching realistic ice thicknesses (Voldoire et al., 2019). We find that the over-insulation in CNRM models may be more related to September–November snow thickness, since the CNRM models show the largest area of 15-30 cm deep snow of

all the models across this time frame (Supplementary Fig. S11), but show similar snow thicknesses compared to other models during December–February (Supplementary Fig. S12). Delayed melt onset could also be related to the use of the GELATO sea ice model, as the CNRM models are the only models used in this study that use the GELATO model (Supplementary Table S2). Since GELATO has a single snow-on-sea-ice layer and fixed albedos for dry snow and melting snow (0.88 and 0.77 respectively) (Voldoire et al., 2019), simplified processes in GELATO may contribute to snow biases.

5 Conclusions

Seasonal sea ice transitions can be characterized by various metrics (melt onset, opening, break-up, freeze onset, freeze-up and closing), and each metric represents a distinct stage of sea ice loss or gain. As such, seasonal transitions provide unique insights into Arctic sea ice processes, but they have so far been under-utilized in evaluating climate models due to a lack of long-term observational products and daily model output, as well as the complexities of defining seasonal transitions. Taking advantage of newly available daily model output (Notz et al., 2016) and observational data of seasonal transitions (Steele et al., 2019), we show that models capture the observed asymmetry in seasonal sea ice transitions, with spring ice loss taking about ~~1.5–2~~ 1-2 months longer than fall ice growth (Figs. 2–7). Models also generally agree with satellite data on the timing of spring transitions~~, but eleven~~ . For fall transition dates, ten out of sixteen models show ~~median~~ mean freeze onset dates later than observed, such that the differences between each model's ~~median~~ mean freeze onset date and the observed date exceed the largest estimations of internal variability (Table 3). Likewise, in ~~almost half of the models (seven~~ five out of sixteen ~~)~~models, the difference between the ~~median~~ mean freeze-up date and the observed date exceed the largest estimations of internal variability. Delayed freeze onset and freeze-up extend simulated melt seasons and open water periods respectively, making the outer ice-free period (the time between ice opening and closing) the only inter-seasonal period in which models consistently agree with satellite observations.

We find that differences in seasonal transitions between models are unlikely due to internal variability alone, and are hence likely a reflection of model differences. Sea ice metrics are each impacted differently by internal variability: models do not agree on a metric most affected, and no single model exhibits the greatest internal variability across all metrics. Despite the uncertainty associated with internal variability, all metrics show pan-Arctic model spreads exceeding even the largest estimations of internal variability in seasonal sea ice transition metrics (Tables 3~~and~~ - 5). The largest standard deviations between ensemble members are seen in the inflow regions ~~of~~ for melt and freeze onset dates (Fig. 8), and this is due to the changing interannual position of the ice edge and the variability of surface temperature.

Because differences in seasonal sea ice transition metrics between models are unlikely due only to internal variability, these metrics can be used for evaluating differences between models in terms of other sea ice characteristics. We show that pan-Arctic relationships between transition metrics and sea ice area and thickness depend on the spatial coverage of the metric (Fig. 9). Out of the six transition dates, melt and freeze onset dates consistently cover the largest area of the Arctic, and they are most closely related to pan-Arctic ice area and mean thickness. Low mean summer ice area delays freeze onset (Table 6), which in turn leads to lower March ice thickness (Table 7). Thinner March ice leads to earlier melt onset and, again, low mean summer

ice area (Table 6). Other relationships between sea ice area and thickness are somewhat discernible using opening and closing dates, but almost indistinguishable using break-up and freeze-up dates (Fig. 9). Since the differences in relationship strengths are seen across definitions that use both surface temperature and ice concentration-based definitions, these differences are more likely related to the spatial coverage of the seasonal sea ice transition dates rather than their defining variables (Tables 6 and 7, Fig. 9). While models tend to show later freeze onset than observed, this offset does not impact the ability of the models to produce the observed relationship between lower summer ice area and later freeze onset.

Finally, we demonstrate how seasonal sea ice transition metrics can provide context to sea ice changes that otherwise lack quantified explanations. We find that CNRM-ESM2-1 and CNRM-CM6-1 exhibit biases in both melt onset (late) and ice thickness (thin) but realistic September sea ice area, exemplifying how seasonal ice transitions can compensate for other unrealistic aspects of the sea ice. Late melt onset helps retain thin ice throughout the summer such that both CNRM models exhibit realistic September sea ice areas for the wrong reasons. Seasonal sea ice transitions metrics therefore provide a process-based constraint on model simulations in addition to the commonly used September and March sea ice areas (Stroeve et al., 2012; Rosenblum and Eisenman, 2017).

To conclude, routinely saved daily sea ice variable output (in particular sea ice concentration and surface temperature) will be critical for using seasonal transitions as a new metric to assess and quantify model uncertainties associated with Arctic sea ice simulations. Since a new observational data product for these seasonal sea ice transition now exists (Steele et al., 2019), seasonal sea ice transition dates should be used routinely in the future to better identify model biases in sea ice evolution as well as the sources of these biases.

Data availability. CMIP6 data are publicly available at the World Climate Research Programme (WCRP) CMIP6, supported by the Department of Energy's Lawrence Livermore National Laboratory and the Earth System Grid Federation (https://esgf-node.llnl.gov/projects/cmip6/, last access, 19 February 2020). All CMIP6 model output used is cited in the reference list. CESM LE data are publicly available at the National Center for Atmospheric Research Climate Data Gateway (https://www.earthsystemgrid.org/, last access: 19 February 2020). Version 1 of the Arctic Sea Ice Seasonal Change and Melt/Freeze Climate Indicators from Satellite Data are publicly available at the National Snow and Ice Data Center (https://nsidc.org/data/NSIDC-0747/versions/1, last access: 26 Aug 2019).

Author contributions. AS and AJ conceived the study, and AS analyzed the data and prepared the manuscript, with guidance and edits from AJ and MW.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We acknowledge the WCRP, which through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. We are grateful for the efforts of SIMIP, supported by the Climate and Cryosphere core project of the WCRP, for requesting the new daily variables used in our analysis and encouraging process-based model analysis. A. Smith's contribution is supported by the Future Investigators in Earth System Science Grant no. 80NSSC19K1324, the National Science Foundation Graduate Research Fellowship grant no. DGE 1144083, and NSF-OPP award 1847398. A. Jahn acknowledges support from NSF-OPP award 1847398. M. Wang is supported by NSF grant 1751363. She is also funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement NA15OAR4320063, Contribution No. 2020-1056 and by NOAA Arctic Research Program, Pacific Marine Environmental Laboratory contribution number 5076. ~~We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF.~~ The CESM project is supported by the National Science Foundation and the Office of Science (BER) of the U.S. Department of Energy. Computing resources for the CESM LE were provided by the Climate Simulation Laboratory at NCAR's Computational and Information Systems Laboratory (CISL), sponsored by the National Science Foundation and other agencies. Five of the CESM LE simulations were produced at the University of Toronto under the supervision of Paul Kushner. NCL (2017) was used for data analysis.

References

Ballinger, T., Lee, C., Sheridan, S., Crawford, A., Overland, J., and Wang, M.: Subseasonal atmospheric regimes and ocean background forcing of Pacific Arctic sea ice melt onset, Climate Dynamics, 52, 5657–5672, https://doi.org/10.1007/s00382-018-4467-x, 2019.

Barnhart, K. R., Miller, C. R., Overeem, I., and Kay, J. E.: Mapping the future expansion of Arctic open water, Nature Climate Change, 6, 1–36, https://doi.org/10.1038/NCLIMATE2848, 2016.

Belchanksy, G., Douglas, D., and Platonov, N.: Duration of the Arctic Sea Ice Melt Season : Regional and Interannual Variability, Journal of Climate, 17, 67–80, https://doi.org/10.1175/1520-0442(2004)017<0067:DOTASI>2.0.CO;2, 2004.

Bitz, C. M. and Roe, G. H.: A Mechanism for the High Rate of Sea Ice Thinning in the Arctic Ocean, Journal of Climate, 17, 3623–3632, https://doi.org/10.1175/1520-0442(2004)017<3623:AMFTHR>2.0.CO;2, 2004.

Bliss, A. and Anderson, M.: Arctic sea ice melt onset from passive microwave satellite data: 1979-2012, The Cryosphere Discussions, 8, 3037–3055, https://doi.org/10.5194/tcd-8-3037-2014, 2014.

Bliss, A., Steele, M., Peng, G., Meier, W., and Dickinson, S.: Regional variability of Arctic sea ice seasonal change climate indicators from a passive microwave climate data record, Environmental Research Letters, 14, 1–11, https://doi.org/10.1088/1748-9326/aafb84, 2019.

Bliss, A. C., Miller, J. A., and Meier, W. N.: Comparison of passive microwave-derived early melt onset records on Arctic sea ice, Remote Sensing, 9, 1–23, https://doi.org/10.3390/rs9030199, 2017.

Boucher, O., Denvil, S., Caubel, A., and Foujols, M.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20190912, https://doi.org/10.22033/ESGF/CMIP6.5195, 2018.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Ethé, C., Fairhead, L., Falletti, L., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, L., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khadre-Traoré, A., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thieblemont, R., Traoré, A., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and evaluation of the IPSL-CM6A-LR climate model, Journal of Advances in Modeling Earth System, https://doi.org/10.1029/2019MS002010, 2020.

Bunzel, F., Notz, D., and Pederson, L.: Retrievals of Arctic Sea-Ice Volume and Its Trend Significantly Affected by Interannual Snow Variability, Geophysical Research Letters, 11, 751–759, https://doi.org/10.1029/2018GL078867, 2018.

Comiso, J., Cavalieri, D., Parkinson, C., and Gloersen, P.: Passive microwave algorithms for sea ice concentration: A comparison of two techniques, Remote Sensing of the Environment, 60, 357–384, https://doi.org/10.1016/S0034-4257(96)00220-9, 1997.

Danabasoglu, G.: NCAR CESM2 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20190912, https://doi.org/10.22033/ESGF/CMIP6.7627, 2019a.

Danabasoglu, G.: NCAR CESM2-FV2 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20200414, https://doi.org/10.22033/ESGF/CMIP6.11297, 2019b.

Danabasoglu, G.: NCAR CESM2-WACCM model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20190912, https://doi.org/10.22033/ESGF/CMIP6.10071, 2019c.

Danabasoglu, G.: NCAR CESM2-WACCM-FV2 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20200414, https://doi.org/10.22033/ESGF/CMIP6.11298, 2019d.

DeRepentigny, P., Jahn, A., Holland, M., and Smith, A.: Arctic Sea Ice in Two Configurations of the Community Earth System Model Version 2 (CESM2) During the 20th and 21st Centuries, Journal of Geophysical Research: Oceans, https://doi.org/10.1029/2020JC016133, 2020.

Dix, M., Bi, D., Dobrohotoff, P., Fiedler, R., Harman, I., Law, R., Mackallah, C., Marsland, S., O'Farrell, S., Rashid, H., Srbinovsky, J., Sullivan, A., Trenham, C., Vohralik, P., Watterson, I., Williams, G., Woodhouse, M., Bodman, R., Dias, F., Domingues, C., Hannah, N., Heerdegen, A., Savita, A., Wales, S., Allen, C., Druken, K., Evans, B., Richards, C., Ridzwan, S., Roberts, D., Smillie, J., Snow, K., Ward, M., and Yang, R.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20200214, https://doi.org/10.22033/ESGF/CMIP6.4271, 2019.

Drobot, S. D. and Anderson, M. R.: An improved method for determining snowmelt onset dates over Arctic sea ice using scanning multichannel microwave radiometer and Special Sensor Microwave/Imager data, Journal of Geophysical Research: Atmospheres, 106, 24,033–24,049, https://doi.org/10.1029/2000JD000171, 2001.

Döscher, R. et al.: The EC-Earth3 Earth System Model for the Climate Model Intercomparison Project 6, in preparation.

EC-Earth-Consortium: EC-Earth-Consortium EC-Earth3 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20200214, https://doi.org/10.22033/ESGF/CMIP6.4700, 2019.

Eyring, V., Bony, S., Meehl, G., Senior, C., Stevens, B., Stouffer, R., and Taylor, K.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A Framework for Collaborative Research, Bulletin of the American Meteorological Society, 94, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1, 2013.

Ivanova, N., Pedersen, L., Tonboe, R., Kern, S., Heygster, G., Lavergne, T., Sørensen, A., Saldo, R., Dybkjær, G., Brucker, L., and Shokr, M.: Inter-comparison and evaluation of sea ice algorithms: towards further identification of challenges and optimal approach using passive microwave observations, The Cryosphere, 9, 1797–1817, https://doi.org/10.5194/tc-9-1797-2015, 2015.

Jahn, A., Sterling, K., Holland, M. M., Kay, J. E., Maslanik, J. A., Bitz, C. M., Bailey, D. A., Stroeve, J., Hunke, E. C., Lipscomb, W. H., and Pollak, D. A.: Late-twentieth-century simulation of Arctic sea ice and ocean properties in the CCSM4, Journal of Climate, 25, 1431–1452, https://doi.org/10.1175/JCLI-D-11-00201.1, 2012.

Johnson, M. and Eicken, H.: Estimating Arctic sea-ice freeze-up and break-up from the satellite record: A comparison of different approaches in the Chukchi and Beaufort Seas, Elementa: Science of the Anthropocene, 4, 1–16, https://doi.org/10.12952/journal.elementa.000124, 2016.

Kashiwase, H., Ohshima, K., Nihashi, S., and Eicken, H.: Evidence for ice-ocean albedo feedback in the Arctic Ocean shifting to a seasonal ice zone, Nature Scientific Reports, 7, 8170, https://doi.org/10.1038/s41598-017-08467-z, 2017.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J. F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., and Vertenstein, M.: The Community Earth System Model (CESM) Large Ensemble project : A community resource for studying climate change in the presence of internal climate variability, Bulletin of the American Meteorological Society, 96, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1, 2015.

Lebrun, M., Vancoppenolle, M., Madec, G., and Massonnet, F.: Arctic sea-ice-free season projected to extend into autumn, The Cryosphere, 13, 79–96, https://doi.org/10.5194/tc-13-79-2019, 2019.

Markus, T., Stroeve, J. C., and Miller, J.: Recent changes in Arctic sea ice melt onset, freezeup, and melt season length, Journal of Geophysical Research: Oceans, 114, 1–14, https://doi.org/10.1029/2009JC005436, 2009.

Mortin, J. and Graversen, R. G.: Evaluation of pan-Arctic melt-freeze onset in CMIP5 climate models and reanalyses using surface observations, Climate Dynamics, 42, 2239–2257, https://doi.org/10.1007/s00382-013-1811-z, 2014.

NCC: NCC NorESM2-LM model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20200214, http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.NCC.NorESM2-LM.historical, 2018a.

NCC: NCC NorESM2-MM model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20200214, http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.NCC.NorESM2-MM.historical, 2018b.

NCL: The NCAR Command Language (Version 6.4.0), Boulder, Colorado: UCAR/NCAR/CISL/TDD , https://doi.org/http://dx.doi.org/10.5065/D6WD3XH5, 2017.

Notz, D., Jahn, A., Holland, M., Hunke, E., Massonnet, F., Stroeve, J., Tremblay, B., , and Vancoppenolle, M.: The CMIP6 Sea-Ice Model Intercomparison Project (SIMIP): understanding sea ice through climate-model simulations, Geoscientific Model Development, 9, 3427–3446, https://doi.org/10.5194/gmd-9-3427-2016, 2016.

O'Neill, B., Tebaldi, C., van Vuuren, D., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G., Moss, R., Riahi, K., and Sanderson, B.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geophysical Research Letters, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

Pegau, W. and Paulson, C.: The albedo of Arctic leads in summer, Annals of Glaciology, 33, 221– 224, https://doi.org/10.3189/172756401781818833, 2001.

Perovich, D.: Sunlight, clouds, sea ice, albedo, and the radiative budget: the umbrella versus the blanket, The Cryosphere, 12, 2159–2165, https://doi.org/10.5194/tc-12-2159-2018, 2018.

Perovich, D., Grenfell, T., Light, B., and Hobbs, P.: Seasonal evolution of the albedo of multiyear Arctic sea ice, Journal of Geophysical Research, 107, C10, https://doi.org/10.1029/2000JC000438, 2002.

Perovich, D., Richter-Menge, J., Jones, K., and Light, B.: Sunlight, water, and ice: Extreme Arctic sea ice melt during the summer of 2007, Geophysical Research Letters, 35, L11 501, https://doi.org/10.1029/2008GL034007, 2008.

Persson, P. O. G.: Onset and end of the summer melt season over sea ice: Thermal structure and surface energy perspective from SHEBA, Climate Dynamics, 39, 1349–1371, https://doi.org/10.1007/s00382-011-1196-9, 2012.

Rosenblum, E. and Eisenman, I.: Sea ice trends in climate models only accurate in runs with biased global warming, Journal of Climate, 30, 6265–6278, https://doi.org/10.1175/JCLI-D-16-0455.1, 2017.

Seferian, R.: CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20190912, https://doi.org/10.22033/ESGF/CMIP6.4068, 2018.

Seland, O., Bentsen, M., Graff, L., Olivié, D., Toniazzo, T., Gjermundsen, A., Debernard, J., Gupta, A., He, Y., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I., Landgren, O., Liakka, J., Moseid, K., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: The Norwegian Earth System Model, NorESM2 - Evaluation of the CMIP6 DECK and historical simulations, Geoscientific Model Development, https://doi.org/10.5194/gmd-2019-378, submitted 2020.

Serreze, M. C., Crawford, A. D., Stroeve, J. C., Barrett, A. P., and Woodgate, R. A.: Variability, trends, and predictability of seasonal sea ice retreat and advance in the Chukchi Sea, Journal of Geophysical Research: Oceans, 121, 7308–7325, https://doi.org/10.1002/2016JC011977, 2016.

SIMIP-Community: Arctic Sea Ice in CMIP6, Geophysical Research Letters, 47, e2019GL086 749, https://doi.org/10.1029/2019GL086749, 2020.

Smith, A. and Jahn, A.: Definition differences and internal variability affect the simulated Arctic sea ice melt season, The Cryosphere, 13, 1–20, https://doi.org/10.5194/tc-13-1-2019, 2019.

Stammerjohn, S., Martinson, D., Smith, R., Yuan, X., and Rind, D.: Trends in Antarctic annual sea ice retreat and advance and their relation to El Nino–Southern Oscillation and Southern Annular Mode variability, Journal of Geophysical Research, 113, 1–20, https://doi.org/10.1029/2007JC004269, 2008.

Stammerjohn, S., Massom, R., Rind, D., and Martinson, D.: Regions of rapid sea ice change: An inter-hemispheric seasonal comparison, Geophysical Research Letters, 39, 1–8, https://doi.org/10.1029/2012GL050874, 2012.

Steele, M., Zhang, J., and Ermold, W.: Mechanisms of summertime upper Arctic Ocean warming and the effect on sea ice melt, Journal of Geophysical Research, 115, C11 004, https://doi.org/10.1029/2009JC005849, 2010.

Steele, M., Dickinson, S., Zhang, J., and Lindsay, R.: Seasonal ice loss in the Beaufort Sea: Toward synchrony and prediction, Journal of Geophysical Research: Oceans, 120, 1118–1132, https://doi.org/10.1002/2014JC010247, 2015.

Steele, M., Bliss, A., Peng, G., Meier, W. N., and Dickinson, S.: Arctic Sea Ice Seasonal Change and Melt/Freeze Climate Indicators from Satellite Data, Version 1, Data subset: 1979-03-01 to 2017-02-28. Data accessed 2019-08-26, 2019.

Stroeve, J. C., Kattsov, V., Barrett, A., Serreze, M., Pavlova, T., Holland, M., and Meier, W. N.: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, Geophysical Research Letters, 39, 1–7, https://doi.org/10.1029/2012GL052676, 2012.

Stroeve, J. C., Markus, T., Boisvert, L., Miller, J., and Barret, A.: Changes in Arctic melt season and implications for sea ice loss, Geophysical Research Letters, 41, 1216–1225, https://doi.org/10.1002/2013GL058951, 2014.

Stroeve, J. C., Crawford, A. D., and Stammerjohn, S.: Using timing of ice retreat to predict timing of fall freeze-up in the Arctic, Geophysical Research Letters, pp. 6332–6340, https://doi.org/10.1002/2016GL069314, 2016.

Swart, N., Cole, J., Kharin, V., Lazare, M., Scinocca, J., Gillett, N., Anstey, J., Arora, V., Christian, J., Hanna, S., Jiao, Y., Lee, W., Majaess, F., Saenko, O., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geoscientific Model Development, 12, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019, 2019a.

Swart, N., Cole, J., Kharin, V., Lazare, M., Scinocca, J., Gillett, N., Anstey, J., Arora, V., Christian, J., Jiao, Y., Lee, W., Majaess, F., Saenko, O., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20190912, https://doi.org/10.22033/ESGF/CMIP6.3610, 2019b.

Timmermans, M. L.: The impact of stored solar heat on Arctic sea ice growth, Geophysical Research Letters, 42, 6399–6406, https://doi.org/10.1002/2015GL064541, 2015.

Voldoire, A.: CMIP6 simulations of the CNRM-CERFACS based on CNRM-CM6-1 model for CMIP experiment historical, Earth System Grid Federation, Version 20190912, https://doi.org/10.22033/ESGF/CMIP6.4066, 2018.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J., Michou, M., Moine, M., Nabat, P., Roehrig, R., Salas y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville,

H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1, Journal of Advances in Modeling Earth Systems, 11, 2177–2213, https://doi.org/10.1029/2019MS001683, 2019.

Walsh, J., Chapman, W., Fetterer, F., and Stewart, J.: Gridded Monthly Sea Ice Extent and Concentration, 1850 Onward, Version 2, Data subset: 1979-01 to 2014-12. Data accessed 2020-01-27, 2019.

Wang, M., Yang, Q., Overland, J. E., and Stabeno, P.: Sea-ice cover timing in the Pacific Arctic: The present and projections to mid-century by selected CMIP5 models, Deep-Sea Research Part II: Topical Studies in Oceanography, 152, 22–34, https://doi.org/10.1016/j.dsr2.2017.11.017, 2018.

Wu, T., Chu, M., Dong, M., Fang, Y., Jie, W., Li, J., Li, W., Liu, Q., Shi, X., Xin, X., Yan, J., Zhang, F., Zhang, J., Zhang, L., and Zhang, Y.: BCC BCC-CSM2MR model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20190912, https://doi.org/10.22033/ESGF/CMIP6.2948, 2018.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., Jie, W., Zhang, J., Liu, Y., Zhang, L., Zhang, F., Zhang, Y., Wu, F., Li, J., Chu, M., Wang, Z., Shi, X., Liu, X., Wei, M., Huang, A., Zhang, Y., and Liu, X.: The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6, Geoscientific Model Development, 12, 1573–1600, https://doi.org/10.5194/gmd-12-1573-2019, 2019.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y., and Ishii, M.: The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2.0: Description and basic evaluation of the physical component, Journal of the Meteorological Society of Japan, p. 1–78, https://doi.org/10.2151/jmsj.2019-051, 2019a.

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yoshimura, H., Shindo, E., Mizuta, R., Ishii, M., Obata, A., and Adachi, Y.: MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20200214, https://doi.org/10.22033/ESGF/CMIP6.6842, 2019b.

Zhang, J., Wu, T., Shi, X., Zhang, F., Li, J., Chu, M., Liu, Q., Yan, J., Ma, Q., and Wei, M.: BCC BCC-ESM1 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, Version 20190912, https://doi.org/10.22033/ESGF/CMIP6.3017, 2018.

Supplementary Material

Citations	Date range	Timing	Threshold	Consecutive Days
Break-up				
Bliss et al. (2019); Steele et al. (2019)	1 March to SIC minimum date	last day	below 15%	–
Serreze et al. (2016)	*X*	first day	below 30%	–
Stammerjohn et al. (2008, 2012)	mid-September to mid-September	last day	below 15%	–
Stroeve et al. (2016)	1 March to SIC minimum date	last day	below 15, 30, 50%	–
Wang et al. (2018)	1 March and 30 September	first day	below 15%	2
Freeze-up				
Bliss et al. (2019); Steele et al. (2019)	SIC minimum date to 28 February	first day	above 15%	–
Serreze et al. (2016)	SIC minimum date to *X*	first day	above 30%	–
Stammerjohn et al. (2008, 2012)	mid-September to mid-September	first day	above 15%	5
Stroeve et al. (2016)	SIC minimum date to 28/29 February	first day	above 15, 30, 50%	–
Wang et al. (2018)	1 September to 31 March	first day	above 15%	2
Open water period				
Barnhart et al. (2016)	11 March to 11 March	number of days	below 15%	–

Table S1. Definitions for break-up (retreat), freeze-up (advance) and the open water period. All studies in the table except Barnhart et al. (2016) calculate the open water period as the number of days between break-up and freeze-up. Information designated with *X* is not provided in the cited manuscripts.

Model	Ocean model	Sea ice model	Ice-ocean Resolution (latitude x longitude)	Cita
ACCESS-CM2	MOM5	CICE5	primarily 1°x 1°	Dix et a
BCC-CSM2-MR	MOM4	SIS2	0.3-1°x 1°	Wu et al. (
BCC-ESM1	MOM4	SIS2	0.3-1°x 1°	Zhang et al. (2018
CanESM5	NEMO3.4.1 ORCA1	LIM2	0.3-1°x 1°	Swart et al
CESM2	POP2	CICE5	0.9°x 1.25°	Danabasog
				DeRepentign
CESM2-FV2	POP2	CICE5	0.9°x 1.25°	Danabasog
CESM2-WACCM	POP2	CICE5	0.9°x 1.25°	Danabasog
				DeRepentign
CESM2-WACCM-FV2	POP2	CICE5	0.9°x 1.25°	Danabasog
CNRM-ESM2-1	NEMO3.6 eORCA1	GELATO6	primarily 1°x 1°	Seferian (2018); V
CNRM-CM6-1	NEMO3.6 eORCA1	GELATO6	primarily 1°x 1°	Voldoire (2018); V
EC-Earth3	NEMO3.6 eORCA1	LIM3	0.3-1°x 1°	~~Döscher et al. (in preparation~~
~~EC-Earth-Consortium (2019)~~IPSL-CM6A-LR	NEMO-OPA eORCA1.3	LIM3	~1°x ~1°	Boucher et al
MRI-ESM2-0	MRI.COM4.4	MRI.COM4.4	0.3-0.5°x 1°	Yukimoto et
NorESM2-LM	MICOM	CICE5	primarily 1°x 1°	NCC (2018a); Seland
NorESM2-MM	MICOM	CICE5	primarily 1°x 1°	NCC (2018b); Seland
CESM LE	POP2	CICE4	0.3-1°x 1°	Hurrell et al. (2013

Table S2. Ocean and sea ice models used by the coupled models, as well as their primary ice-ocean resolutions and associated citations.

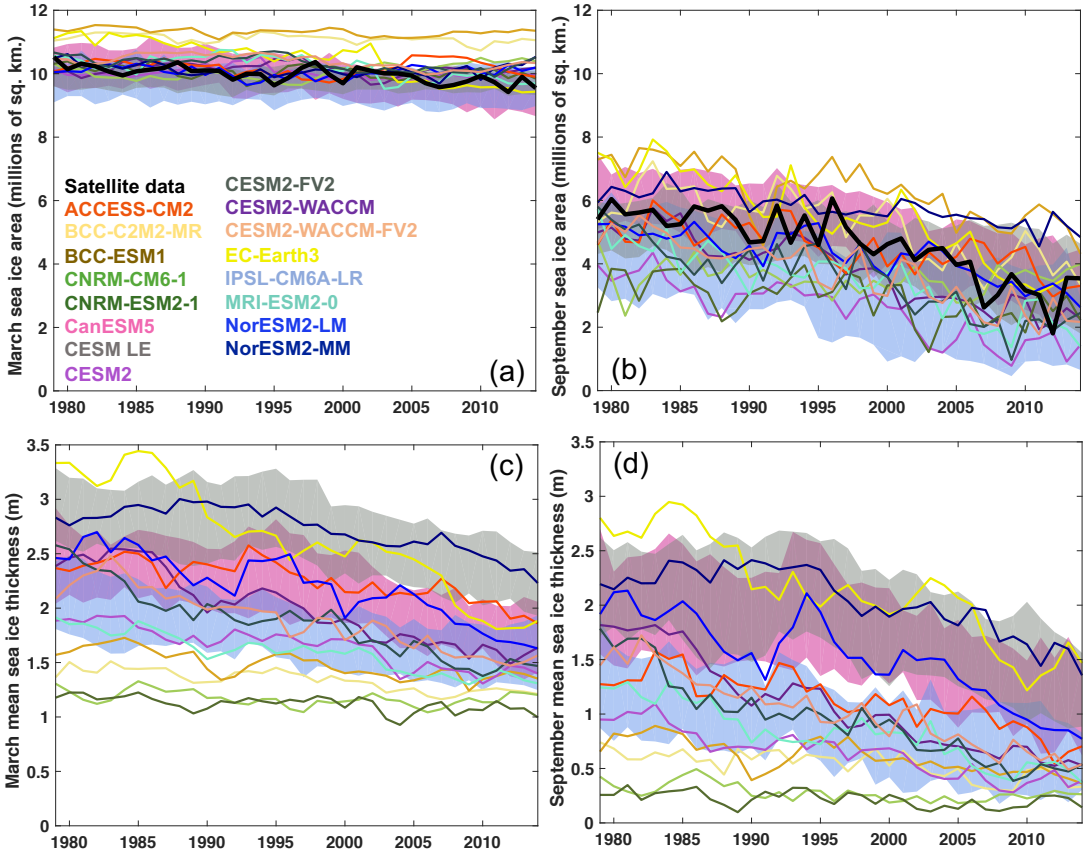

Figure S1. From 1979–2014, (a) March sea ice area (b) September sea ice area (c) March mean ice thickness and (d) September mean ice thickness in CMIP6 models (various colors), CESM LE (gray) and satellite observations (black) in the Arctic. All ensemble members are shown for CESM (40 members), CanESM5 (35 members) and IPSL (30 members). ~~Observations~~ Metrics are averaged from 66-84.5°N and observations of sea ice thickness are not shown.

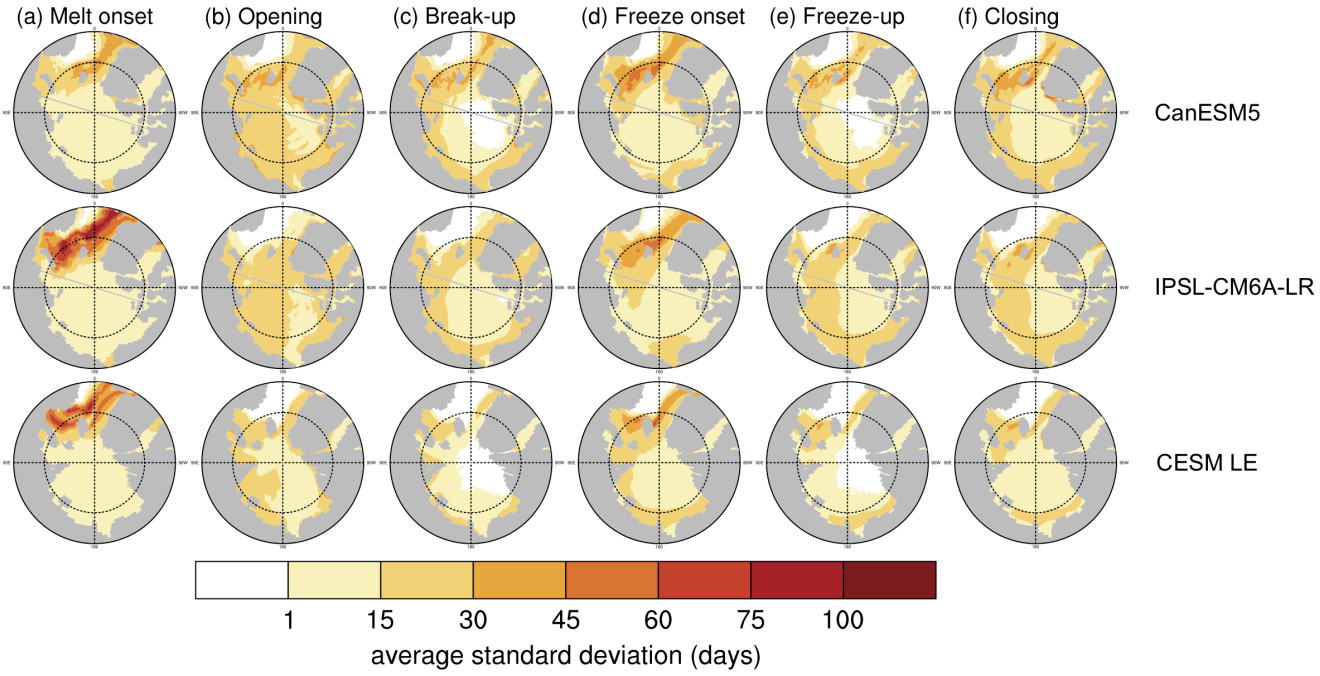

Figure S2. The average standard deviation between all available ensemble members over 1979–2014 for (a) melt onset (b) opening (c) break-up (d) freeze onset (e) freeze-up (f) closing. CanESM5 is displayed in the first row (35 members), IPSL is displayed in the second row (30 members) and CESM LE is displayed in the third row (40 members). The standard deviation is calculated at each grid cell for each year, and then the average of all years is plotted for each grid cell. The same figure using the first 30 ensemble members of each model is displayed in Fig. 8.

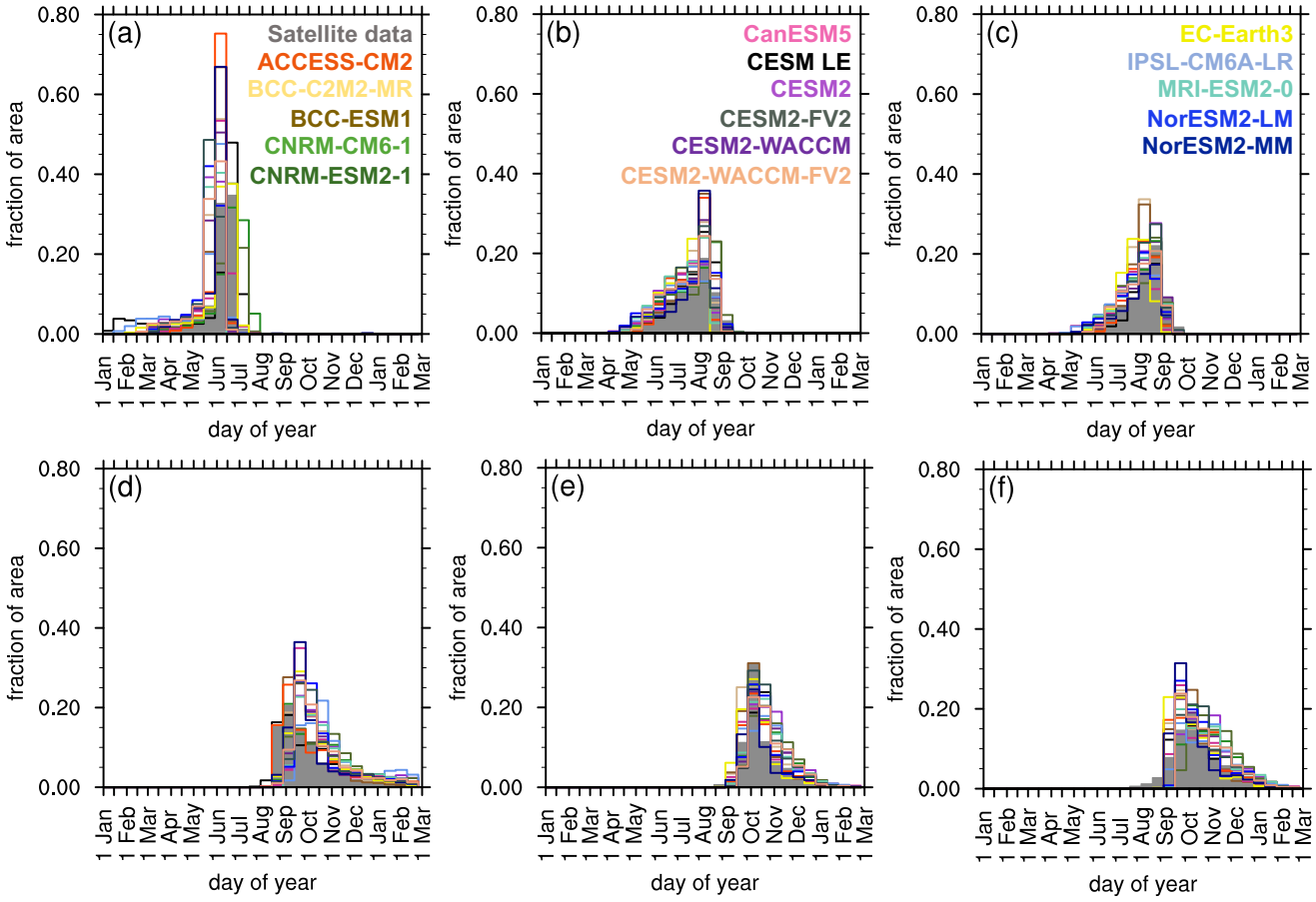

Figure S3. Area distributions of the average of each metric from 1979–2014: (a) melt onset (b) opening (c) break-up (d) freeze onset (e) freeze-up and (f) closing. Metrics are averaged from 66-84.5°N for satellite data (filled gray) and the first ensemble member of each model (all other colors). All models and satellite data are represented in each panel (a)-(f), but the color labels are distributed across panels (a)-(c).

	Melt onset	Opening (80%)	Break-up (15%)	Freeze onset	Freeze-up (15%)	Closing (80%)
ACCESS-CM2	~~0.52~~[a] 12	~~0.74~~[a] 24	19	48	24	30
BCC-CSM2-MR	18	25	18	31	22	30
BCC-ESM1	17	24	18	30	21	27
CanESM5	18	28	26	41	29	35
CESM2	22	27	23	42	26	32
CESM2-FV2	18	26	22	41	24	29
CESM2-WACCM	21	27	23	40	25	31
CESM2-WACCM-FV2	21	28	24	39	25	31
CNRM-ESM2-1	24	32	28	48	26	30
CNRM-CM6-1	25	31	27	47	26	30
EC-Earth3	29	23	20	36	22	28
IPSL-CM6A-LR	36	31	29	43	29	34
MRI-ESM2-0	19	28	26	42	26	32
NorESM2-LM	20	34	27	39	24	30
NorESM2-MM	22	32	27	39	26	30
CESM LE	47	26	19	48	23	31
Satellite data	20	32	27	44	27	36

Table S3. Satellite-era (1979–2014) spatial standard deviations across the Arctic (in days, between 66-84.5°N) of seasonal sea ice transition dates calculated using the satellite data and the first ensemble member from each model.

	Melt period	Seasonal loss-of-ice period	Freeze period	Seasonal gain-of-ice period	Melt season	Open water period	Outer ice-free period
ACCESS-CM2	18	9	8	7	49	52	41
BCC-CSM2-MR	17	12	7	7	47	53	39
BCC-ESM1	20	13	7	7	44	49	37
CanESM5	16	10	7	6	49	61	54
CESM2	21	8	7	7	59	56	46
CESM2-FV2	21	10	6	6	54	52	43
CESM2-WACCM	22	10	6	8	57	55	45
CESM2-WACCM-FV2	22	10	7	6	57	57	46
CNRM-ESM2-1	21	7	6	6	63	60	52
CNRM-CM6-1	18	8	8	5	63	60	50
EC-Earth3	17	10	12	8	59	51	40
IPSL-CM6A-LR	22	8	15	5	71	62	55
MRI-ESM2-0	20	11	7	7	56	60	49
NorESM2-LM	24	10	9	6	54	61	48
NorESM2-MM	19	14	9	7	52	61	51
CESM LE	16	12	14	10	83	39	56
Satellite data	21	12	29	10	58	49	61

Table S4. Satellite-era (1979–2014) spatial standard deviations across the Arctic (in days, between 66-84.5°N) of intra-seasonal periods (melt period, seasonal loss-of-ice period, freeze period and seasonal gain-of-ice period) and inter-seasonal periods (melt season, open water period and outer ice-free period) calculated using the satellite data and the first ensemble member from each model.

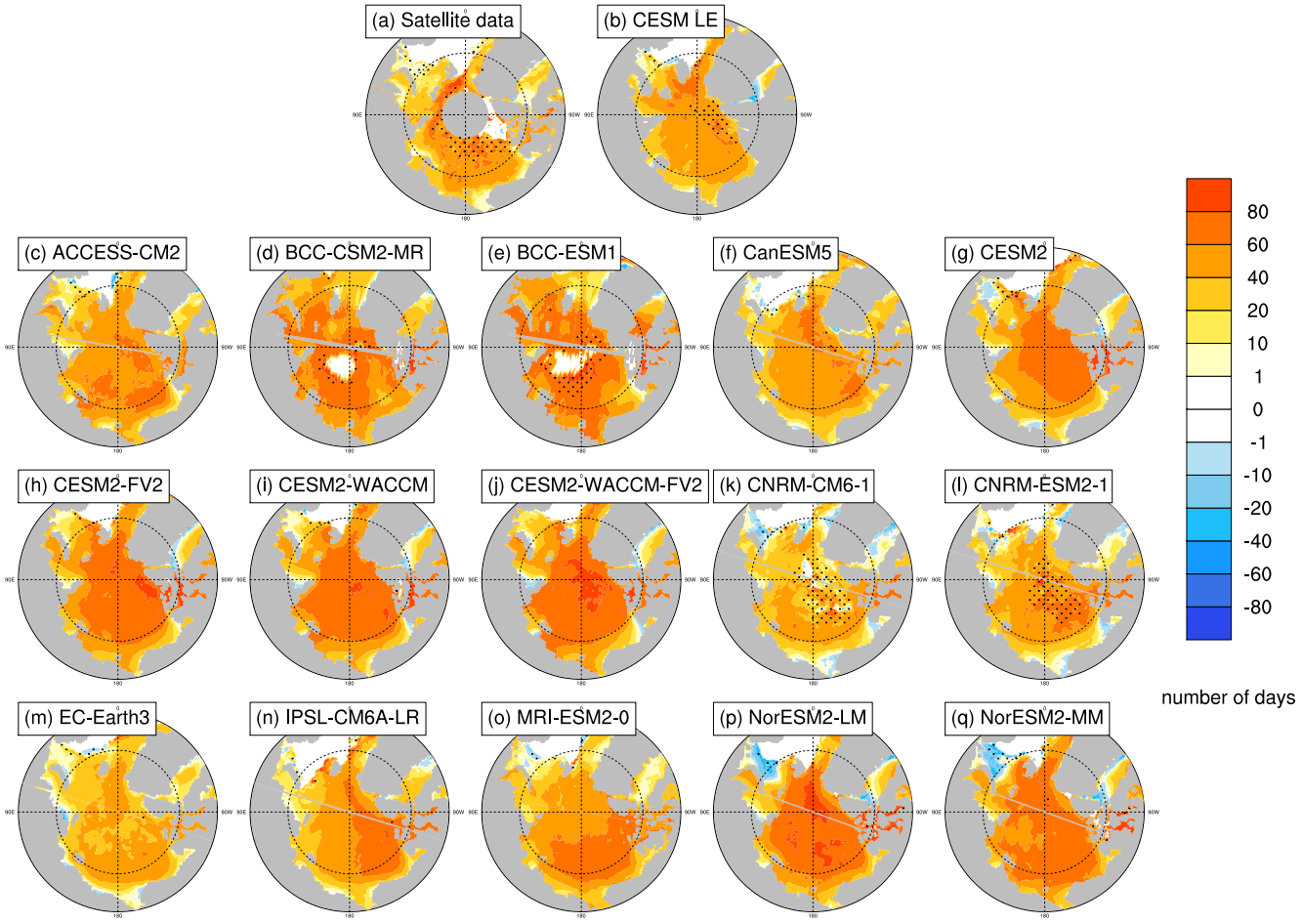

Figure S4. The length of the melt period (number of days between melt onset and opening) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where closing dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines. Negative values indicate where the opening date falls earlier than the melt onset. This can occur due to physical reasons (i.e., dynamical ice divergence or bottom melt), or due to the fact that melt onset is defined using surface temperature and opening is defined using ice concentration.

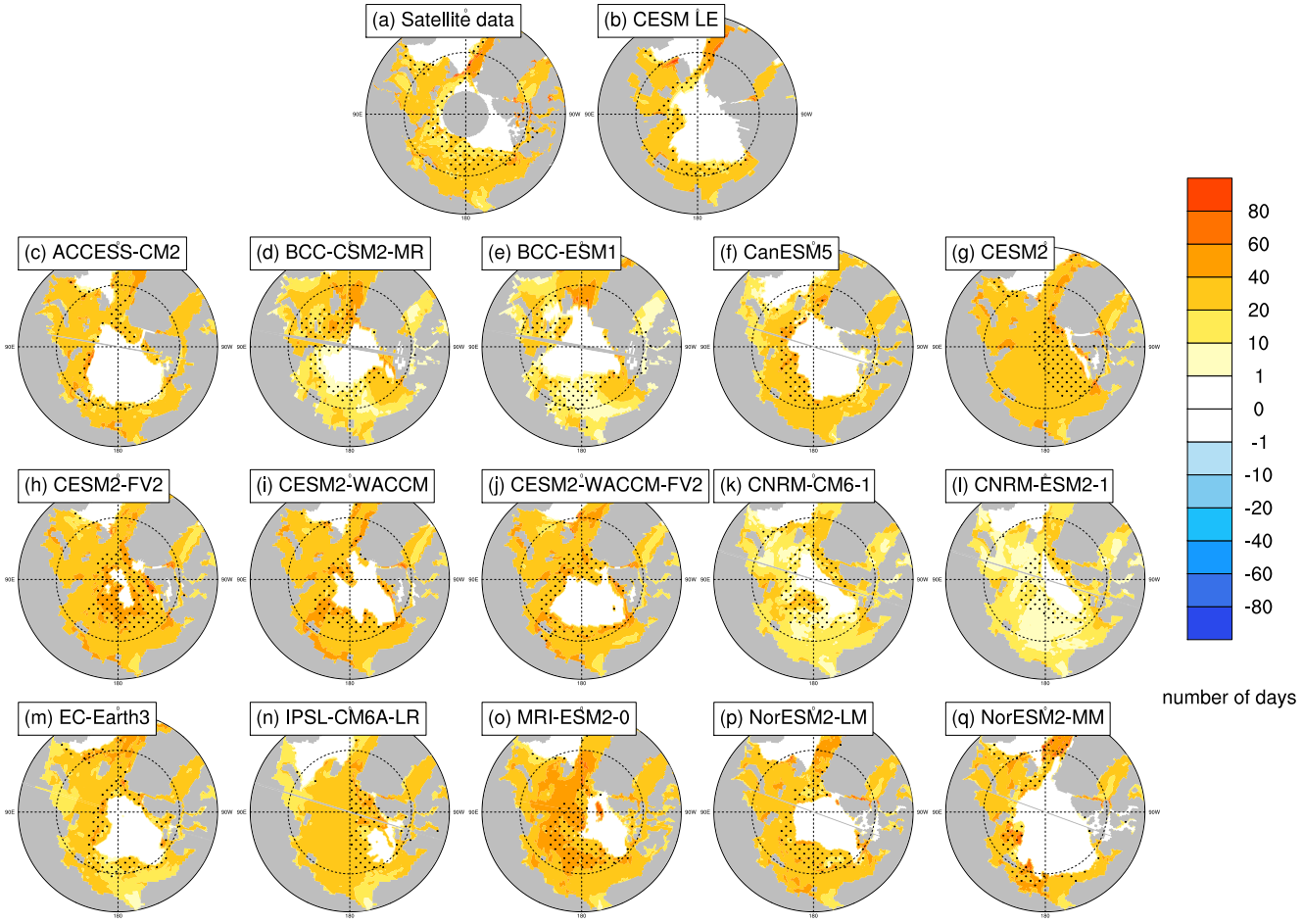

Figure S5. The length of the seasonal loss-of-ice period (number of days between opening and break-up) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where closing dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines. No negative values are possible as both metrics are based on sequential ice concentration thresholds (80% and 15%).

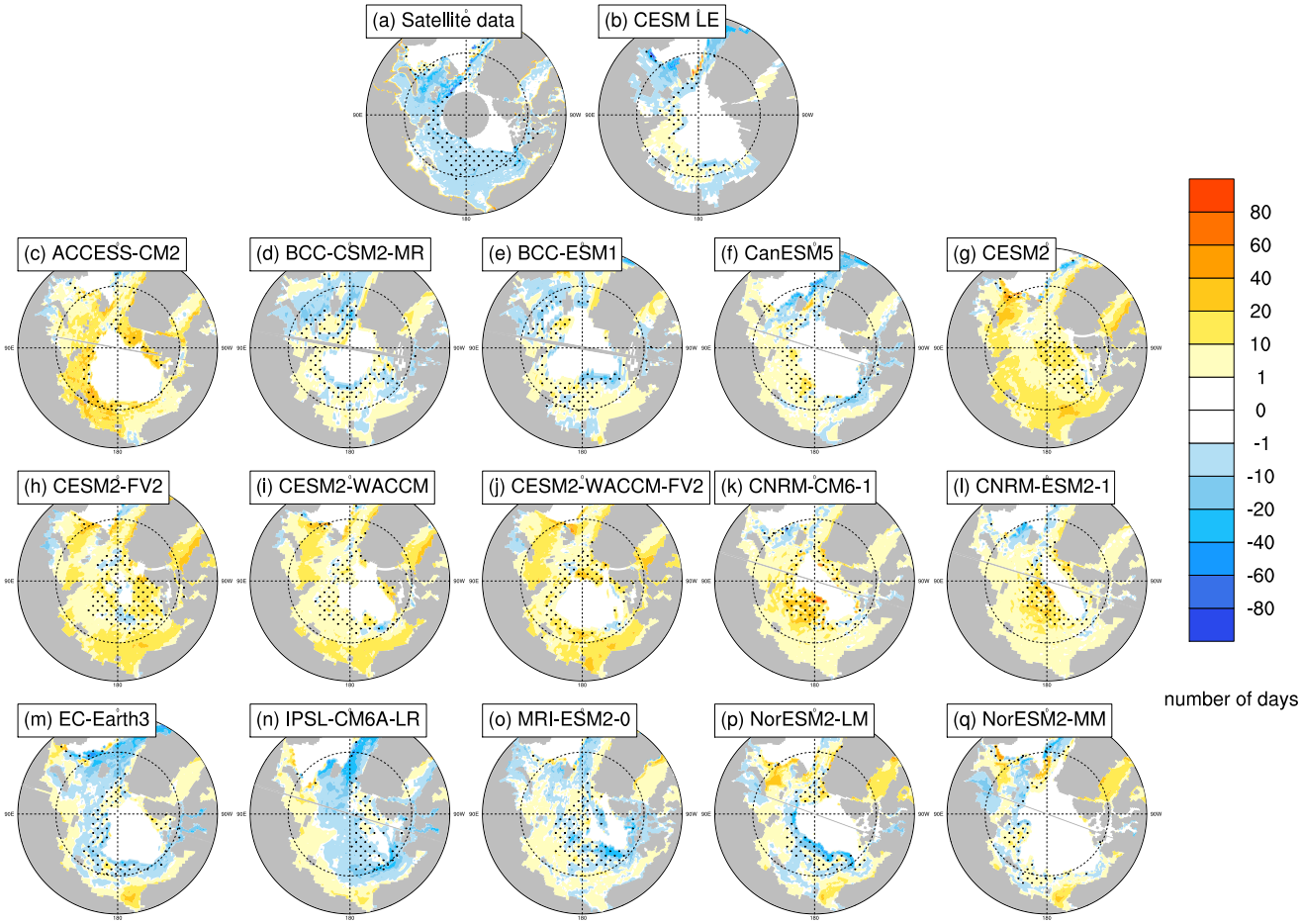

Figure S6. The length of the freeze period (number of days between freeze onset and freeze-up) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where closing dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines. Negative values indicate where the freeze-up date falls earlier than the freeze onset. This can occur due to physical reasons (i.e., dynamical ice convergence), or due to the fact that freeze onset is defined using surface temperature and freeze-up is defined using ice concentration.

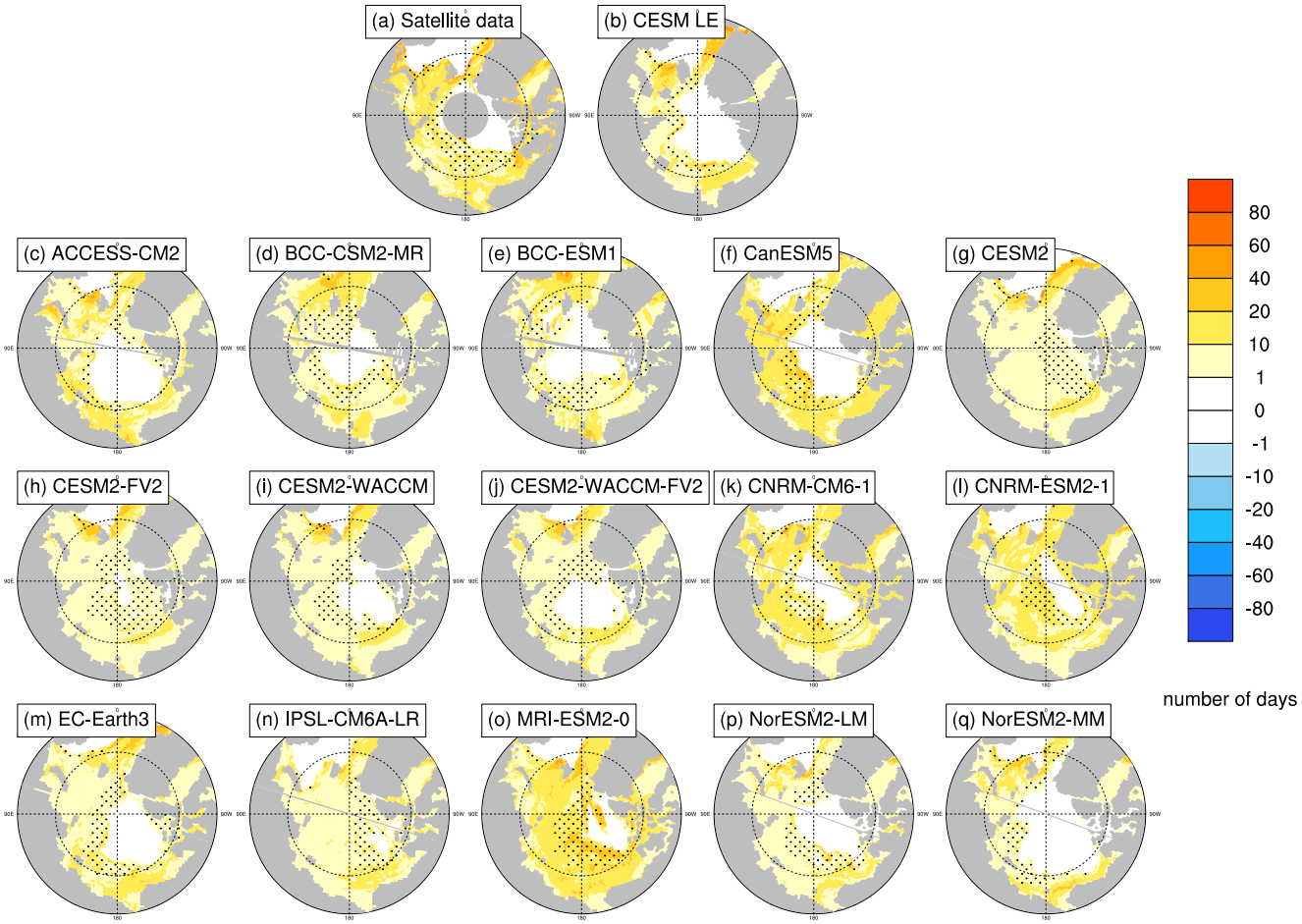

Figure S7. The length of the seasonal gain-of-ice period (number of days between freeze-up and closing) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where closing dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines. No negative values are possible as both metrics are based on sequential ice concentration thresholds (15% and 80%).

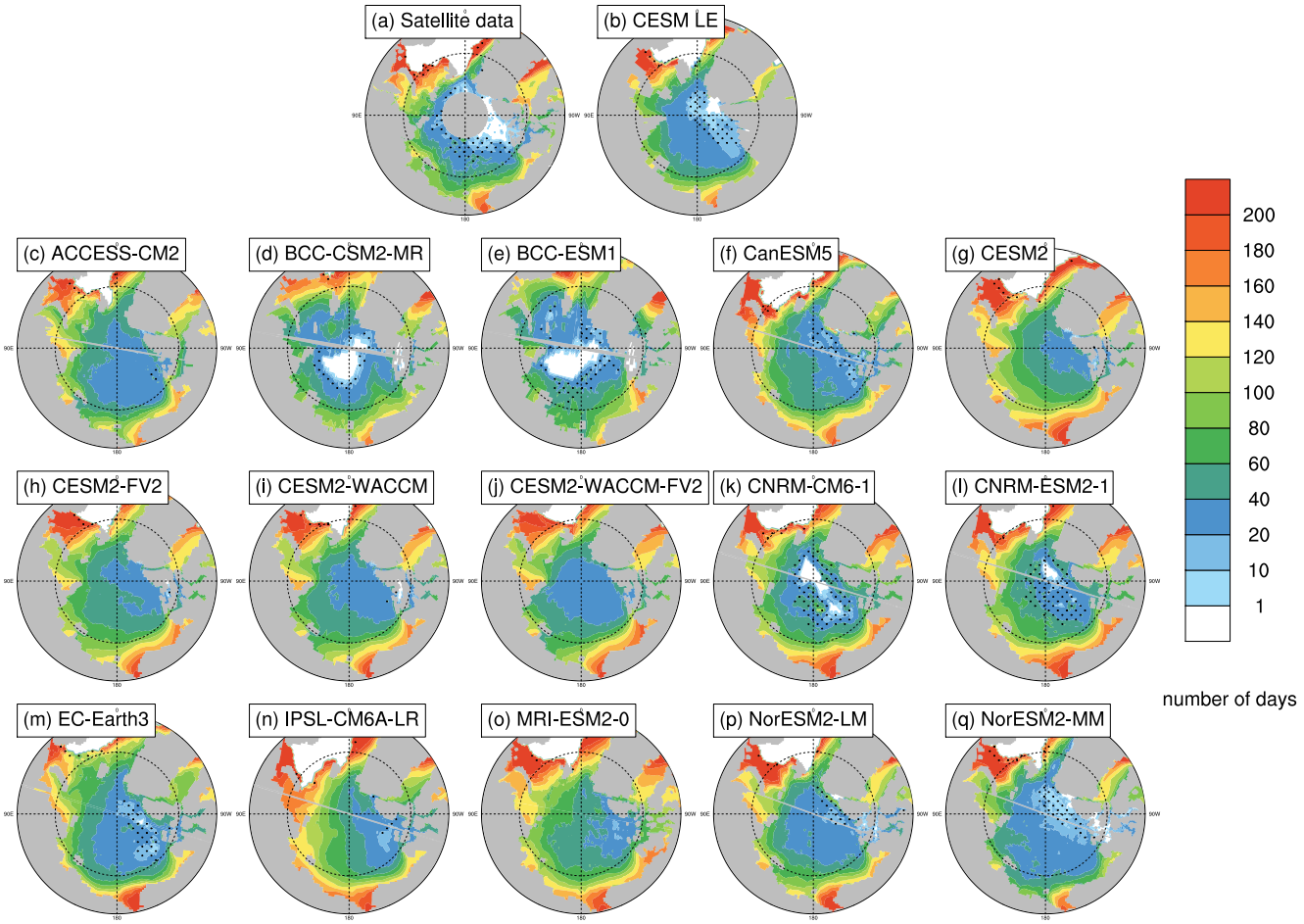

Figure S8. The length of the outer ice-free period (number of days between opening and closing, which use an 80% ice concentration threshold) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where closing dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines.

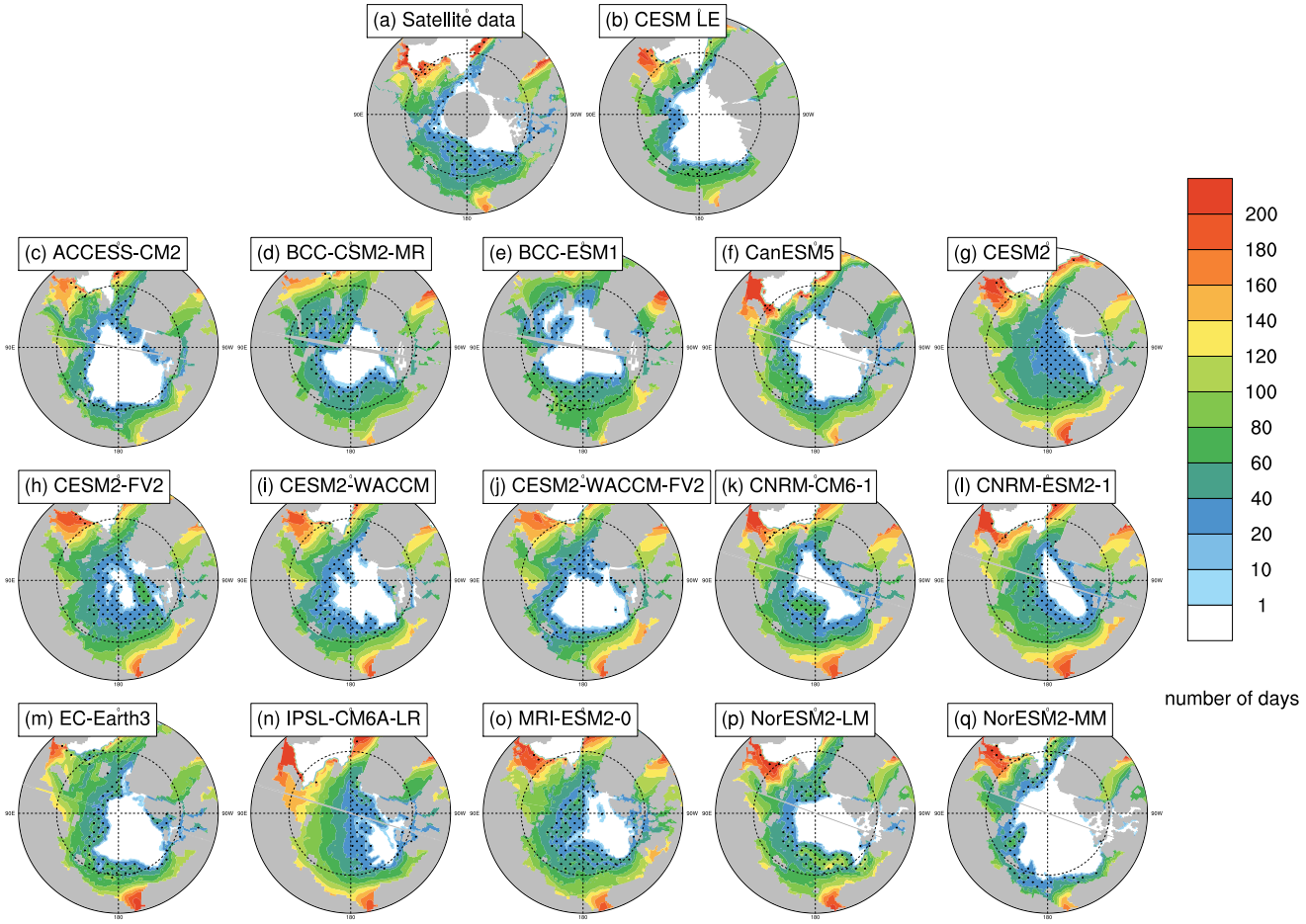

Figure S9. The length of the open water period (number of days between break-up and freeze-up, which use a 15% ice concentration threshold) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where closing dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines.

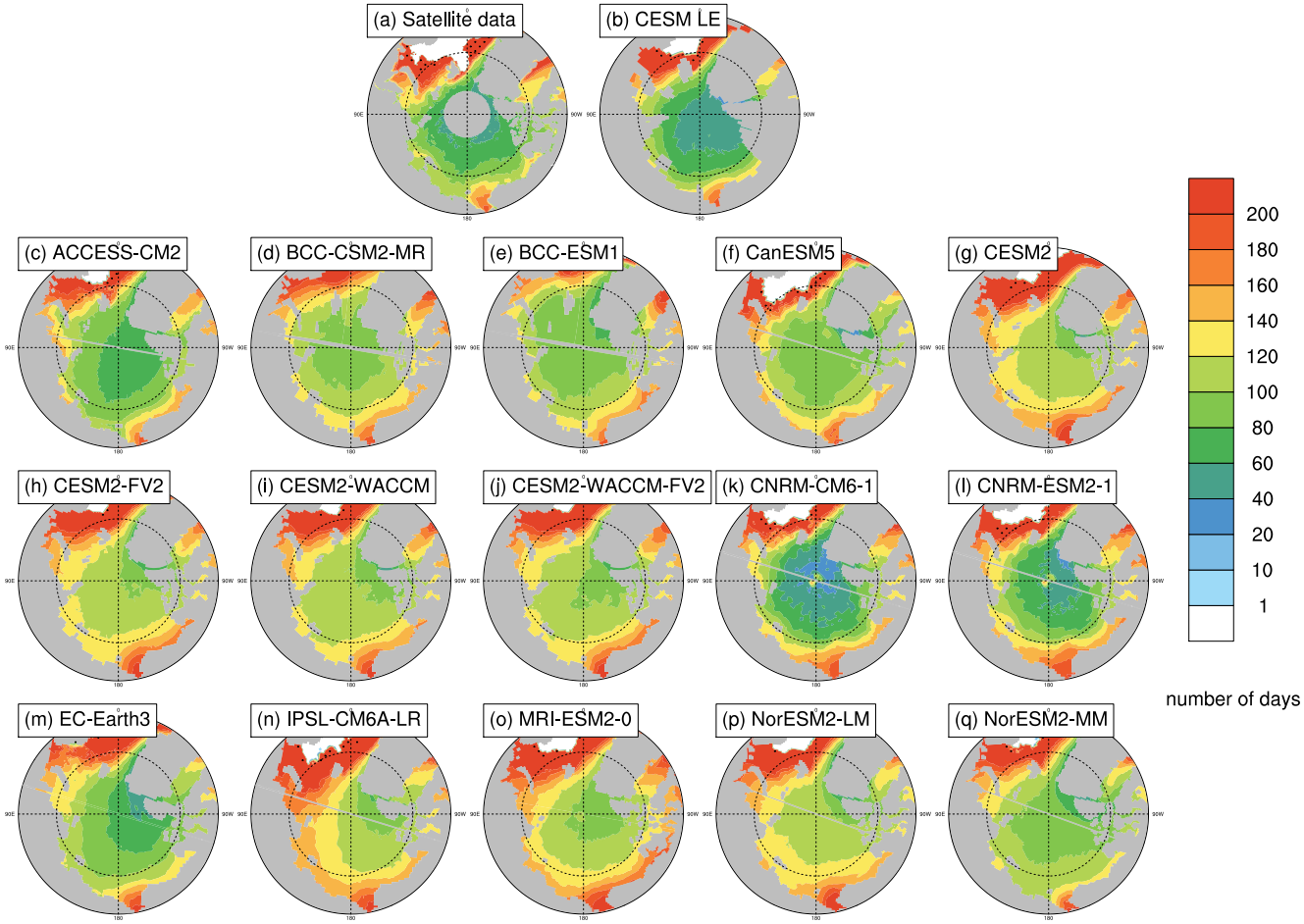

Figure S10. The length of the melt season (number of days between melt onset and freeze onset, which are defined using surface temperature in the models and brightness temperatures in the satellite data) averaged over 1979–2014 at each grid cell using satellite data (a), the first ensemble member of the CESM LE (b) and the first ensemble member of each CMIP6 model (c-q). Stippling indicates where closing dates exist in less than 20% of years in the time range. Models on tripolar grids produce plot gaps filled by gray lines.

	Melt onset	Opening (80%)	Break-up (15%)	Freeze onset	Freeze-up (15%)	Closing (80%)
ACCESS-CM2	**0.52**	**0.74**	0.24	~~-0.76~~[a] **-0.76**	~~-0.74~~[a] **-0.74**	~~-0.67~~[a] **-0.67**
BCC-CSM2-MR	0.33	~~0.46~~[a] **0.46**	0.16	~~-0.70~~[a] **-0.70**	~~-0.47~~[a] **-0.47**	~~-0.66~~[a] **-0.66**
BCC-ESM1	~~0.55~~[a] **0.55**	-0.16	0.05	~~-0.69~~[a] **-0.69**	~~-0.58~~[a] **-0.58**	~~-0.47~~[a] **-0.47**
CESM2	~~0.38~~[a] **0.38**	~~0.76~~[a] **0.76**	0.30	~~-0.82~~[a] **-0.82**	~~-0.75~~[a] **-0.75**	~~-0.86~~[a] **-0.86**
CESM2-FV2	~~0.66~~[a] **0.66**	~~0.73~~[a] **0.73**	0.09	~~-0.85~~[a] **-0.85**	~~-0.69~~[a] **-0.69**	~~-0.74~~[a] **-0.74**
CESM2-WACCM	~~0.53~~[a] **0.53**	~~0.65~~[a] **0.65**	-0.07	~~-0.73~~[a] **-0.73**	~~-0.48~~[a] **-0.48**	~~-0.67~~[a] **-0.67**
CESM2-WACCM-FV2	~~0.49~~[a] **0.49**	~~0.67~~[a] **0.67**	~~0.36~~[a] **0.36**	~~-0.79~~[a] **-0.79**	~~-0.67~~[a] **-0.67**	~~-0.73~~[a] **-0.73**
CNRM-ESM2-1	~~0.46~~[a] **0.46**	-0.15	-0.11	-0.19	-0.13	-0.17
CNRM-CM6-1	0.19	-0.18	-0.11	-0.07	-0.04	-0.08
EC-Earth3	~~0.75~~[a] **0.75**	~~0.53~~[a] **0.53**	~~0.48~~[a] **0.48**	~~-0.91~~[a] **-0.91**	~~-0.82~~[a] **-0.82**	~~-0.79~~[a] **-0.79**
MRI-ESM2-0	~~0.37~~[a] **0.37**	~~0.46~~[a] **0.46**	-0.18	~~-0.83~~[a] **-0.83**	~~-0.76~~[a] **-0.76**	~~-0.79~~[a] **-0.79**
NorESM2-LM	~~0.47~~[a] **0.47**	~~0.63~~[a] **0.63**	0.03	~~-0.68~~[a] **-0.68**	~~-0.46~~[a] **-0.46**	~~-0.61~~[a] **-0.61**
NorESM2-MM	~~0.44~~[a] **0.44**	0.08	~~-0.47~~[a] **-0.47**	~~-0.53~~[a] **-0.53**	~~-0.4~~[a] **-0.4**	~~-0.38~~[a] **-0.38**
CanESM5	~~0.67~~[a] **0.68**	~~0.49~~[a] **0.50**	~~-0.12~~[a] **-0.11**	~~-0.6~~[a] **-0.61**	~~-0.53~~[a] **-0.54**	~~-0.57~~[a] **-0.57**
IPSL-CM6A-LR	~~0.41~~[a] **0.42**	~~0.49~~[a] **0.49**	~~-0.08~~[a] **-0.08**	~~-0.78~~[a] **-0.78**	~~-0.66~~[a] **-0.66**	~~-0.71~~[a] **-0.71**
CESM LE	~~0.35~~[a] **0.35**	~~0.19~~[a] **0.19**	~~-0.11~~[a] **-0.11**	~~-0.76~~[a] **-0.76**	~~-0.34~~[a] **-0.35**	~~-0.43~~[a] **-0.43**

Table S5. ~~Correlation~~ As in Table 6, correlation coefficients (R-values) between seasonal sea ice transition dates ~~and~~ but with summer (June–September) mean sea ice thickness of the same year instead of sea ice area from 1979–2014. Values ~~with~~ [a] in bold are statistically significant at the 95% confidence level. Correlation coefficients and p-values for models in the first thirteen rows are determined using one ensemble member, for CanESM5 using all 35 ensemble members, for IPSL using all 30 ensemble members and CESM LE using all 40 ensemble members. All values are calculated between 66-84.5°N.

	Melt onset	Opening (80%)	Break-up (15%)	Freeze onset	Freeze-up (15%)	Closing (80%)
ACCESS-CM2	-0.16	0.06	-0.15	-0.13 0.00	-0.06 -0.04	-0.03 -0.06
BCC-CSM2-MR	0.49[a] **0.49**	0.25	0.29	-0.34[a] **-0.41**	-0.21 **-0.35**	-0.26 **-0.43**
BCC-ESM1	0.29	-0.17	-0.19	-0.45[a] **-0.36**	-0.27 **-0.36**	-0.21 **-0.43**
CESM2	0.06	0.38[a] **0.38**	0.15	-0.43[a] -0.19	-0.32 -0.22	-0.39[a] **-0.25**
CESM2-FV2	0.44[a] **0.44**	0.51[a] **0.51**	0.14	-0.79[a] **-0.69**	-0.61[a] **-0.61**	-0.61[a] **-0.60**
CESM2-WACCM	0.21	0.42[a] **0.42**	0.41[a] **0.41**	-0.25 **-0.39**	-0.20 **-0.50**	-0.22 **-0.50**
CESM2-WACCM-FV2	0.49[a] **0.49**	0.44[a] **0.44**	0.22	-0.68[a] **-0.64**	-0.44[a] **-0.49**	-0.48[a] **-0.52**
CNRM-ESM2-1	-0.51[a] **-0.51**	-0.49[a] **-0.49**	-0.46[a] **-0.46**	0.70[a] **0.59**	0.58[a] **0.53**	0.62[a] **0.56**
CNRM-CM6-1	-0.09	-0.09	-0.28	-0.26 -0.11	0.14 0.06	-0.13 -0.08
EC-Earth3	0.79[a] **0.79**	0.56[a] **0.56**	0.44[a] **0.44**	-0.85[a] **-0.84**	-0.72[a] **-0.77**	-0.73[a] **-0.75**
MRI-ESM2-0	0.29	0.43[a] **0.43**	0.08	-0.66[a] **-0.53**	-0.54[a] **-0.59**	-0.52[a] **-0.53**
NorESM2-LM	-0.02	-0.07	0.18	-0.06 0.18	0.19 0.03	0.13 0.05
NorESM2-MM	0.32	-0.2	-0.05	-0.27 -0.19	-0.14 **-0.35**	0.00 -0.12
CanESM5	0.49[a] **0.50**	0.39[a] **0.39**	0.04	-0.66[a] **-0.64**	-0.42[a] **-0.54**	-0.43[a] **-0.52**
IPSL-CM6A-LR	0.39[a] **0.40**	0.36[a] **0.36**	0.18[a] **0.19**	-0.48[a] **-0.39**	-0.40[a] **-0.44**	-0.38[a] **-0.40**
CESM LE	0.12[a] **0.12**	-0.02	-0.03 -0.04	-0.11[a] **-0.07**	0.07[a] **0.13**	0.06[a] **0.10**
Satellite data	0.56[a] **0.57**	0.42[a] **0.42**	0.39[a] **0.38**	-0.58[a] **-0.61**	-0.27 **-0.35**	-0.44[a] **-0.48**

Table S6. ~~Correlation~~ As in Table 7, correlation coefficients (R-values) between seasonal sea ice transition dates ~~and~~ but with March sea ice area instead of sea ice thickness from 1979–2014. Spring transition dates (melt onset, opening and break-up) are correlated with March mean ice area from the same year, while fall transition dates (freeze onset, freeze-up and closing) are correlated with March mean ice area from the following year. Values ~~with [a]~~ in bold are statistically significant at the 95% confidence level. Correlation coefficients and p-values for models in the first thirteen rows are determined using one ensemble member, for CanESM5 using all 35 ensemble members, for IPSL using all 30 ensemble members and CESM LE using all 40 ensemble members. All values are calculated between 66-84.5°N.

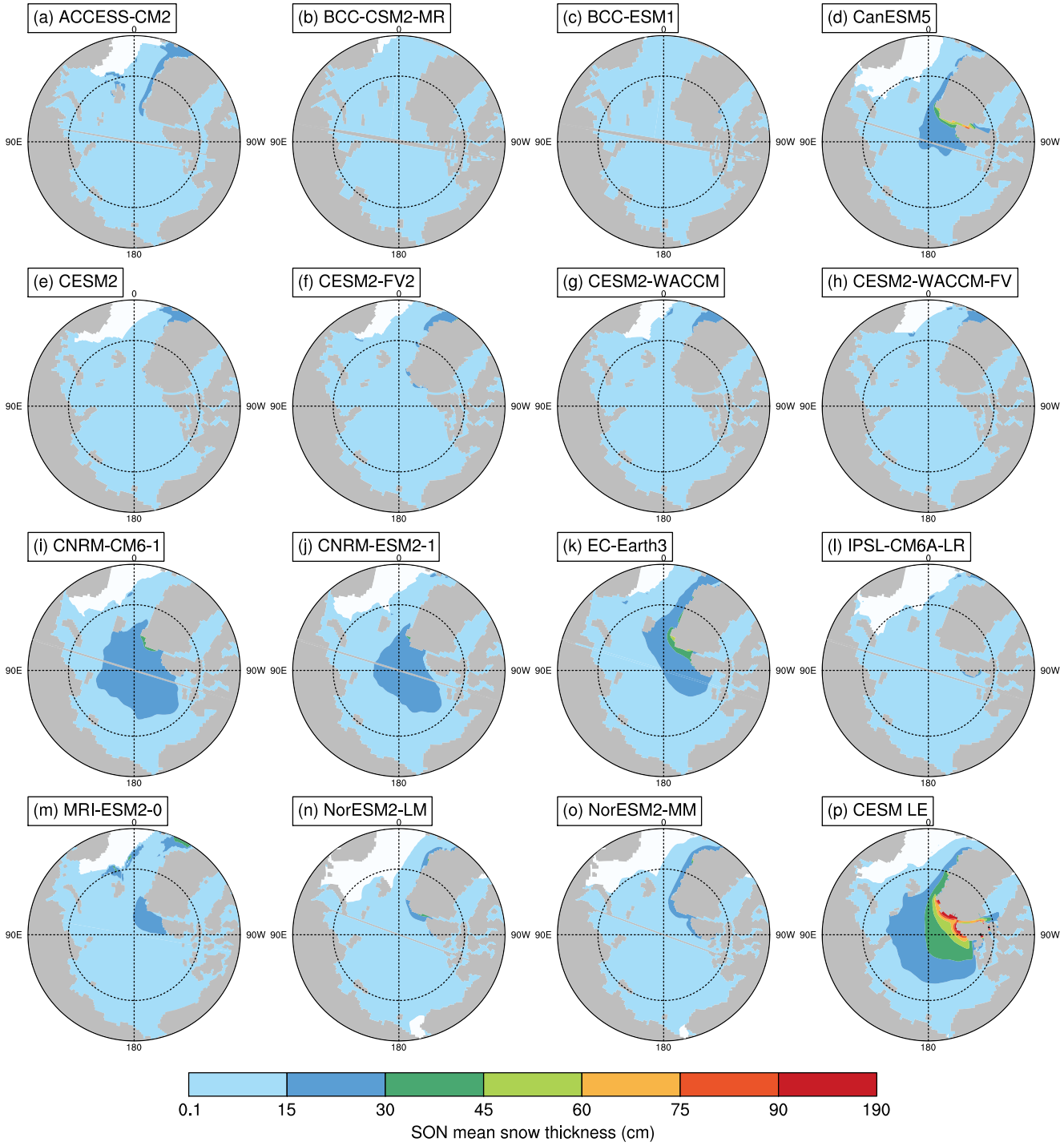

Figure S11. September–November mean snow thickness using the first ensemble member of each CMIP6 model (a-o) and the CESM LE (p) from 1979–2014. Note that the largest contour interval ~~is~~ spans 100 cm instead of 15 cm to account for the very high snow depths in the CESM LE.

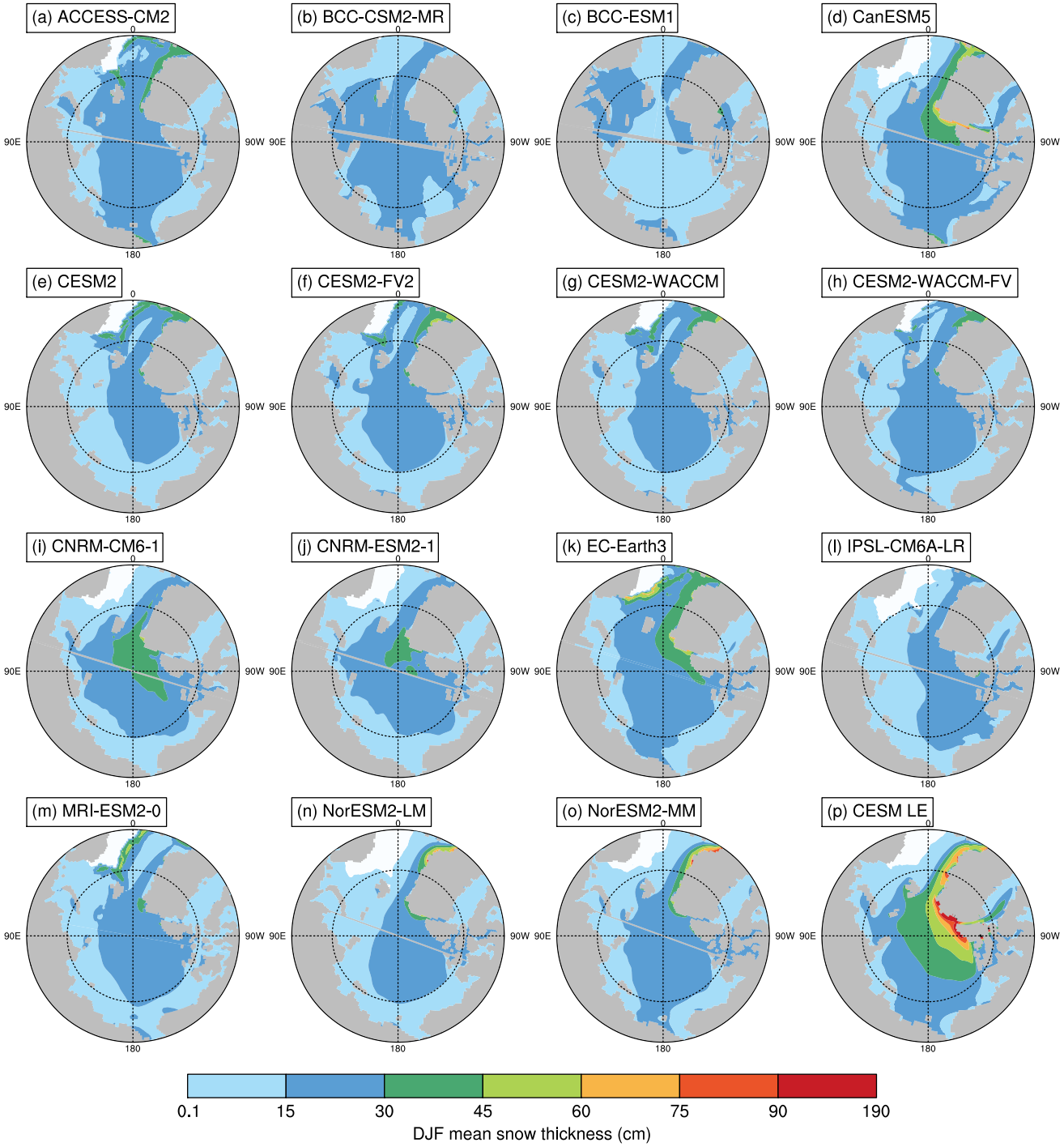

Figure S12. December–February mean snow thickness using the first ensemble member of each CMIP6 model (a-o) and the CESM LE (p) from 1979-2014. Note that the largest contour interval ~~is~~ spans 100 cm instead of 15 cm to account for the very high snow depths in the CESM LE.