On the representation of Arctic sea ice in climate models

Author: Daniel Garcia Pinazo*
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: María Santolaria-Otín

Abstract: The performance of seven models participating in the 6th phase of the Coupled Model Intercomparison Project (CMIP6) in simulating historical Arctic sea ice concentration is assessed against observations. Although individually the models are not too accurate, the CMIP6 multi-model mean adequately replicates the mean annual cycle and the declining sea ice observed trends. Regional model biases have also been identified in the spatial distribution of sea ice.

I. INTRODUCTION

Sea ice in the Arctic plays a crucial role for the local people and the wildlife that depend upon it. Not only that, but its importance stretches far beyond, to the point of affecting our entire climate system. Arctic sea ice regulates Earth's energy balance, reflecting sunlight back to outer space thanks to its white surface (high albedo), and prevents heat exchange between the atmosphere above and the ocean below [1].

The Arctic is a very vulnerable region to global warming. It warms at a pace more than twice as fast as the global average, a phenomenon known as Arctic amplification. Consequently, rises in surface air temperature in response to increases in atmospheric greenhouse gas concentrations are larger in the poles than in the rest of the world [2]. Due to the sea ice loss, the Arctic undergoes a feedback loop that threatens to culminate in an ice-free Arctic Ocean.

Climate models help us to understand the Earth's system and can be used to estimate the consequences of further changes in our climate. The combined analysis of observations and model simulations gives a strong insight into the past and the future evolution of Arctic sea ice [3]. The Coupled Model Intercomparison Project (CMIP) of the World Climate Research Programme (WCRP) began in 1995 and is now in its sixth phase (CMIP6). CMIP aims to coordinate the global climate modelling community and provide a multi-model perspective of the climate system. In this study, we will evaluate the performance of seven CMIP6 models in simulating Arctic sea ice during the present period by comparing the data from the historical experiment of the models with observational data.

Our goal is to estimate how accurate the models are and identify any shortcomings each model might have. By comparing the models with observational data, we can find where they fail to replicate the past sea ice state, so we can consider the bias to correct the future model predictions of sea ice evolution.

*Electronic address: dgarcipi9@alumnes.ub.edu

II. DATA AND METHODOLOGY

A. Data

The variables considered for the characterization of sea ice in this study are the sea ice concentration (SIC), the sea ice extent (SIE), and the sea ice area (SIA). The SIC (%) is defined as the percentage of a grid cell covered by ice and has been used for the spatial analysis of the sea ice. The SIE (km^2) is computed as the sum of the areas of all grid cells with at least 15% SIC [4]. The SIA (km^2) is computed as the sum of the ice concentration times the area of all grid cells with at least 15% SIC. According to these definitions, the SIE will always be larger than the SIA. As a result, the SIE reduces the uncertainties in SIC and SIA summertime values taken from satellite sensors, which tend to be underestimated due to the sensors identifying surface melt as open water instead of water on top of sea ice [4]. For this reason, the SIE has been chosen as the variable for the temporal analysis of sea ice. Finally, the sea ice edge is defined as having a SIC of 15% and indicates the boundary between ice and open water.

The observational data is the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST) [5]. The HadISST provides the monthly fields of SIC on a 1° latitude-longitude grid from 1870 to date. We have also used monthly SIC from historical runs (1850 to 2014) of seven models from the Coupled Model Intercomparison Project phase 6 (CMIP6). From each model, we have subjectively selected only one ensemble member. The period considered for both observations and models is January 1979 to December 2014. The starting date has been chosen because, from 1979, the observational data has been obtained by satellites, making it more reliable. The region we have considered as the Arctic ranges from 60° latitude to 90°.

The climate models used are EC-Earth3 [6], CanESM5 [7], CESM2 [8], IPSL-CM6A-LR [9], MIROC6 [10], NorESM-LM [11], and UKESM1-0-LL [12]. The models have been chosen to show a broad spread in sea ice.

B. Methodology

As each model has a different spatial resolution in a curvilinear grid, in order to compare the observations with the models, we have interpolated the model data sets to the observations grid using the nearest source-to-destination method from xESMF [13].

The trends have been calculated using a linear regression. A Mann-Kendall test with the null hypothesis of no-monotonic trend has been used to evaluate the statistical significance of the trends. When the null hypothesis is rejected, we conclude that a significant trend exists. The multi-model mean (MM) has been computed as an average of the seven individual models and its uncertainty as an average of individual standard deviations.

The climatology has been calculated as the mean of SIC for the 1979-2014 period at the grid point level. Spatial bias has been calculated as model climatology minus observations climatology. A two-sample t-test with the null hypothesis of equal means has been used to evaluate the statistical significance of the bias. When the null hypothesis is rejected, the difference between the model and the observations climatology is concluded as statistically significant and not caused by random error. Spatial standard deviation has been computed using the standard deviation of the SIC values in every point along the temporal coordinate. A f-test with the null hypothesis of equal variances has been used to evaluate the statistical significance of the spatial standard deviation. When the null hypothesis is rejected, the difference between the model variance and the observed variance is concluded to be statistically significant. All statistical tests have been carried out at a 95% confidence level.

In order to evaluate a model's ability to replicate the observations, the root-mean-square-error (RMSE) and the spatial correlation coefficient (r) have been calculated for each model. The RMSE has been calculated by doing the square root of the spatial average of the square deviation of a model's climatology from the observed one. The closer the RMSE is to zero, the better the model replicates the observations. The r between each model climatology and the observations climatology has been calculated using the Python function *corrcoef()* from the NumPy package. The closer the r is to 1, the better the model replicates the observations

III. RESULTS

First, we evaluate the state of Arctic sea ice in the observations from the HadISST data set. In Figure 1a, SIE trends are negative for all months except for May, when the trend is positive but very small. Only April, May, and June show non-significant trends. August, September, October, and November display the highest trends, with September and October having trends of around $-60 \times 10^3 km^2 yr^{-1}$. In Figure 1b, these same months also present the most marked anomalies, and two different trends can be distinguished, one that ranges until about the year 2001 (period P1, from 1979 to 2001) and one that goes onward from then and is stronger and negative (period P2, from 2002 to 2014). Some years in the last period have remarkably high anomalies. The year 2012 stands out, with September showing an anomaly of $-2.8 \times 10^6 km^2$. September 2012 remains the lowest minimum Arctic sea ice extent registered to date. The years 2007 and 2011 are notable as well. We can also see how the interannual variability, the change from year to year, is much higher in summer and early fall than in winter and spring.

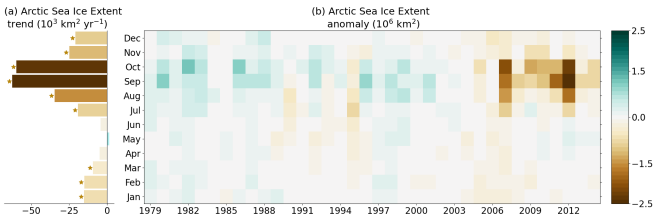

FIG. 1: Monthly Arctic SIE (a) trends and (b) anomalies of the HadISST data set. Stars in trends mark them as significant, using a Mann-Kendall test at a 95% confidence level.

Figure 2 displays the Arctic SIE annual cycle in models and HadISST. The observations present a very pronounced minimum of $6.5 \times 10^6 km^2$ in September and a not-so-clear maximum of $13.4 \times 10^6 km^2$ in March. The amplitude, defined as the difference between the maximum and minimum SIE, of the observed annual cycle is $6.9 \times 10^6 km^2$. The CMIP6 MM is very close to the observations. It replicates the observed minimum and the beginning of the growing season (September, October, and November) almost identically, while it underestimates the SIE during the rest of the year. As a result, the amplitude of the annual cycle is smaller; for example, the March maximum is slightly shifted down by $0.2 \times 10^6 km^2$. However, the underestimation is never drastic, being at most of $0.5 \times 10^6 km^2$ during July, so the observations fall within the range of the MM standard deviation for all months. The MM standard deviation is highest during August, September, and October. We can relate this high values to the two different trends of the periods P1 and P2 in Figure 1.

Individual models spread widely above and under observations. All models have the maximum in March, and only EC-Earth3 has the minimum in August instead of September. MIROC6 yields the smallest amplitude, $5.1 \times 10^6 km^2$, but gets the closest to observations in the minimum of the annual cycle. CanESM5 overestimates SIE almost all months, but the shape of its annual cycle is similar to the observations resulting in an amplitude closest to the observed one, $7.0 \times 10^6 km^2$. IPSL-CM6A-LR and CESM2 simulate the closest SIE to observations during winter but underestimate it at the minimum by around $1.6 \times 10^6 km^2$, the closest simulations to an ice-free ocean in summer.

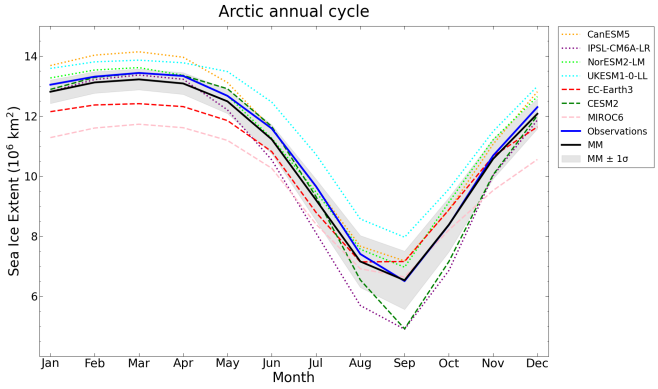

FIG. 2: Arctic SIE annual cycle of the CMIP6 models and the HadISST observations. The CMIP6 multi-model mean (MM) is in black and its one standard deviation uncertainty range is in grey shading.

Figure 3 shows the annual cycle of monthly trends for observations and the ensemble of models. The trend is negative for all months except May; it is highest in September and is only non-significant in April, May, and June. The CMIP6 MM has been computed without EC-Earth3, as it is an outlier for all months. The MM simulates a higher declining trend than observed for all months except for February when the trend is slightly smaller than observed but also the closest. It is furthest from observations during the melting season, reaching a maximum difference of about $-20 \times 10^3 km^2$, and approaches the observations again at the September negative trend maximum. The error bars for the observations and the MM always overlap and are longest during summer and fall. Again, this overly large values can be attributed to the two different trends of the periods P1 and P2 seen in Figure 1.

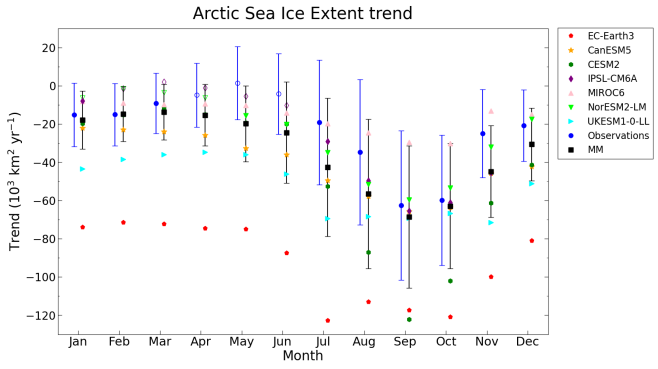

FIG. 3: Monthly Arctic SIE trends of the CMIP6 models and HadISST. Non-significant trends, using a Mann-Kendall test at a 95% confidence level, are represented with an empty marker. The one standard deviation uncertainty range of the observations is represented by the blue error bars. The CMIP6 multi-model mean (MM) is in black, and its one standard deviation uncertainty range is in black error bars.

The spread of the individual model trends is quite large, displaying values above and below observations. EC-Earth3 is the least accurate at simulating the observed SIE trend, with a difference of over $-60 \times 10^3 km^2$ for most months. UKESM1-0-LL also overestimates the SIE negative trend for all months but does fall within the error bars during summer. CanESM5 also overestimates but falls within the error bars for all months. CESM2 is close to observations in January, February, and March but does not replicate the trend accurately during summer and fall. MIROC6 performs well overall but simulates too much sea ice during summer. IPSL-CM6A-LR and NorESM2-LM show similar good results and are the only models with non-significant trends for some months during winter, spring, and early summer.

For a complete evaluation of the performance of CMIP6 models, the temporal analysis is not enough, as regional biases could be cancelling each other out such that the simulated SIE resembles the observations despite large regional differences [3]. For this reason, the spatial distribution of SIC is presented. We chose to analyze the month of September because it is when the SIE is minimum (Fig. 2) and also when there is the largest trend and anomalies among all months (Figs. 1 and 3).

Figure 4a displays the SIC climatology of the observations. We see the SIC is highest ($\geq 90\%$) in the central region of the Arctic Ocean up to the northern coast of Greenland and the northernmost islands of the Canadian Arctic Archipelago. In this region, the sea is almost totally covered by ice. Then, the SIC decreases gradually in the other directions from the centre. The other panels in Figure 4 show the bias in each model. The observed climatology and the model bias are comparable in magnitude, so the same scale has been used for both. All seven CMIP6 models replicate the SIC distribution with greater or lesser accuracy (see also Fig. 5 in the Appendix).

We can distinguish three groups of models according to the distribution of their September spatial bias in Figure 4. Firstly, EC-Earth3 and UKESM1-0-LL exhibit a homogeneously positive bias, meaning that, in some regions, they simulate more sea ice than there is. These two models simulate larger areas of high ($\sim 70\%$) SIC, and their sea ice edge happens further from the centre. This results in two large, shared regions of significant positive bias ($\sim 50\%$). The first goes through the Fram Strait and the Greenland, Barents, and Kara seas; the second is in the East Siberian, Chukchi, and Beaufort seas. As these outer contours of SIC are where the presence of sea ice changes the most from year to year, it is also to be expected that this region will have a higher standard deviation (see also Fig. 6 in the Appendix). This seems to indicate a relationship exists between high bias and high standard deviation.

CESM2 and IPSL-CM6A form the second group, characterized instead by a negative bias in most of the Arctic region. IPSL-CM6A simulates a high ($\sim -70\%$) deficit of SIC between the Kara and Laptev seas that decreases gradually towards the Arctic Ocean and the East

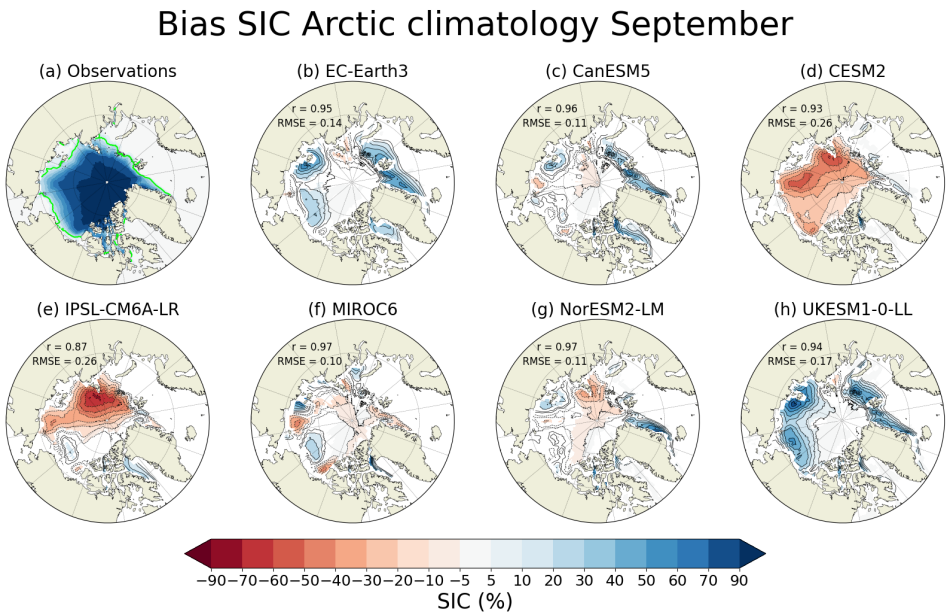

FIG. 4: (a) Arctic September SIC climatology of HadISST. The lime contour represents the sea ice edge, defined as having a SIC of 15%, and indicating the boundary between open water and ice. (b)-(h) Arctic September SIC bias of the CMIP6 models. Black contours represent the total bias, and only statistically significant differences, using a t-test test at a 95% confidence level, are shaded. The RMSE and r values for individual models against observations are included in the upper-left corner of each panel.

Siberian Sea. This bias occurs because the sea ice edge of this model encircles the smallest area out of the seven models. As a result, there is a high ($\sim -60\%$) negative bias just outside the sea ice edge, where the observations exhibit SIC. CESM2 presents a high ($\sim -60\%$) negative bias in the same region as IPSL-CM6A but also underestimates ($\sim -30\%$) the SIC in the Arctic Ocean and the northern coasts of Alaska, Canada, and Greenland. This last bias appears because CESM2 simulates lower SIC than 90% in the region of maximum ($\geq 90\%$) observed SIC, unlike the rest of the models.

Finally, CanESM5, NorESM2-LM, and MIROC6 display a heterogeneous bias, positive in some regions and negative in others. CanESM5 yields a similar bias distribution to EC-Earth3 and UKESM1-0-LL, as seen in the positive bias ($\sim 50\%$) in the Fram Strait and the Greenland, Barents, and Kara seas. However, the extension of the region totally covered by sea ice is smaller and mostly limited to the Arctic Ocean north of the coasts of Greenland and the Canadian Arctic Archipelago, and a low negative bias ($\sim -10\%$) appears in the other half of the Arctic Ocean. NorESM2-LM and MIROC6 positive and negative biases are even more alternately distributed. NorESM2-LM high ($\geq 90\%$) SIC region is also confined to the coasts of Greenland and the Canadian Arctic Archipelago, again causing a negative bias to appear in the Arctic Ocean. MIROC6 has the high SIC region shifted in the direction of the Beaufort Sea, causing a negative bias to appear where the sea should be covered by ice, according to the observations. However, these biases are never too high ($\sim \pm 20\%$). Some biases are shared by all seven models, a positive bias in the Baffin Bay and a negative bias between the Kara and Laptev seas, both regions with a high density of islands.

The RMSE and the r values depicted also give an idea of what models are best at replicating the spatial distribution of SIC of the observations. In the models with a homogeneously negative bias, the bias is high and covers a large region. Because of this, these models have the RMSE value furthest from zero (0.26) and the r value furthest from 1 (0.87 and 0.93). In the models with a homogeneously positive bias, the bias is also high but mostly limited to the contours of the sea ice edge, resulting in better values of RMSE (0.17 and 0.14) and r (0.94 and 0.95). The models with heterogeneous bias do not have their significant bias restricted to the sea ice edge, but their bias is generally lower. The resulting RMSE values are the closest to zero (0.11 and 0.10), and the r values are the closest to 1 (0.96 and 0.97).

IV. CONCLUSIONS

The goal of this study is to evaluate the performance of seven CMIP6 models in simulating sea ice in the Arctic during the period 1979-2014. The initial analysis of the HadISST observational data has corroborated the Arctic sea ice loss, in the shape of high declining trends and negative anomalies of SIE, especially pronounced in recent years.

One of the limitations of this study is that we have only

used one ensemble member for each model. Even then, and although individually the models are not completely accurate, the resulting CMIP6 MM replicates the observations to a remarkable extent. The MM agrees with the observations regarding the strong declining trend of SIE, especially high during summer and September, when SIE is minimum.

From the spatial analysis of the climatology of the CMIP6 models, we have identified several problematic regions for each model and we have grouped the models according to these regions: first models with a homogeneous positive bias (EC-Earth3 and UKESM1-0-LL), second models with a homogeneous negative bias (CESM2 and IPSL-CM6A), and finally, models with a heterogeneous bias (CanESM5, NorESM2-LM, and MIROC6). A more thorough analysis of each model would have to be conducted to ascertain what causes the different biases of SIC in each case.

We have identified a relation between bias and standard deviation. The region with the highest standard deviation is usually the sea ice edge and its surroundings, as this exterior contour is what changes the most between years. Here is where we usually find a higher bias, suggesting that models may have difficulties in simulating the high interannual variability of the sea ice. The models also seem to present difficulties in regions where the percentage of land in the grid is high, such as coastal regions and archipelagos.

The RMSE and r values are computed as indicators for comparing the models' accuracies in replicating the spatial distribution of SIC. According to these parameters, MIROC6 and NorESM5 appear to be the best models.

Our study highlights that not a single model has a generally good performance; instead, each has its strengths and weaknesses. Therefore, we can conclude that there is no better model and that only by considering them together we can recreate reality with enough accuracy to reach solid conclusions. A further study could consist of evaluating whether adding more members to a model would or would not improve its results, in order to see how representative of an individual model performance our results are, and to see if the CMIP6 MM would improve.

Acknowledgments

I want to thank María Santolaria and Mario Rodrigo for their kind advice and guidance through my stay with the Meteorology Group of the University of Barcelona and the elaboration of this study.

[1] Qi Shu, Qiang Wang, Zhenya Song, Fangli Qiao, Jiechen Zhao, Min Chu, and Xinfang Li. Assessment of sea ice extent in cmip6 with comparison to observations and cmip5. *Geophysical Research Letters*, 47, 2020.

[2] M. C. Serreze, A. P. Barrett, J. C. Stroeve, D. N. Kindig, and M. M. Holland. The emergence of surface-based arctic amplification. *Cryosphere*, 3, 2009.

[3] Julienne Stroeve and Dirk Notz. Insights on past and future sea-ice evolution from combining observations and models, 2015.

[4] Matthew Watts, Wieslaw Maslowski, Younjoo J. Lee, Jaclyn Clement Kinney, and Robert Osinski. A spatial evaluation of arctic sea ice and regional limitations in cmip6 historical simulations. *Journal of Climate*, 34, 2021.

[5] N. A. Rayner, D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, 108, 2003.

[6] EC-Earth Consortium (EC-Earth). Ec-earth-consortium ec-earth3 model output prepared for cmip6 cmip historical, 2019.

[7] N. C. Swart, J. N. Cole, V. V. Kharin, M. Lazare, J. F. Scinocca, N. P. Gillett, J. Anstey, V. Arora, J. R. Christian, Y. Jiao, W. G. Lee, F. Majaess, O. A. Saenko, C. Seiler, C. Seinen, A. Shao, L. Solheim, K. von Salzen, D. Yang, B. Winter, and M. Sigmond. Cccma canesm5 model output prepared for cmip6 cmip historical. *Earth Syst.Grid Fed.*, 2019.

[8] Gokhan Danabasoglu, David Lawrence, Keith Lindsay, William Lipscomb, and Gary Strand. Ncar cesm2 model output prepared for cmip6 cmip historical. *Earth System Grid Federation*, 20190912, 2019.

[9] Olivier Boucher, Sébastien Denvil, Guillaume Levavasseur, Anne Cozic, Arnaud Caubel, Marie-Alice Foujols, Yann Meurdesoif, Patricia Cadule, Marion Devilliers, Josefine Ghattas, Nicolas Lebas, Thibaut Lurton, Lidia Mellul, Ionela Musat, and Frédérique Migno. Ipsl ipsl-cm6a-lr model output prepared for cmip6 cmip historical, 2018.

[10] Hiroaki Tatebe and Masahiro Watanabe. Miroc miroc6 model output prepared for cmip6 cmip historical, 2018.

[11] Øyvind Seland, Mats Bentsen, Dirk Jan Leo Oliviè, Thomas Toniazzo, Ada Gjermundsen, Lise Seland Graff, Jens Boldingh Debernard, Alok Kumar Gupta, Yanchun He, Alf Kirkevåg, Jörg Schwinger, Jerry Tjiputra, Kjetil Schanke Aas, Ingo Bethke, Yuanchao Fan, Jan Griesfeller, Alf Grini, Chuncheng Guo, Mehmet Ilicak, Inger Helene Hafsahl Karset, Oskar Andreas Landgren, Johan Liakka, Kine Onsum Moseid, Aleksi Nummelin, Clemens Spensberger, Hui Tang, Zhongshi Zhang, Christoph Heinze, Trond Iversen, and Michael Schulz. Ncc noresm2-lm model output prepared for cmip6 cmip historical. version 20190815. *Earth System Grid Federation*, 2019.

[12] Yongming Tang, Steve Rumbold, Rich Ellis, Douglas Kelley, Jane Mulcahy, Alistair Sellar, Jeremy Walton, and Colin Jones. Mohc ukesm1.0-ll model output prepared for cmip6 cmip historical, 2019.

[13] J Zhuang, R Dussin, A Jüling, and S Rasp. xesmf: Universal regridder for geospatial data. *Zenodo*, 10, 2018.

V. APPENDIX

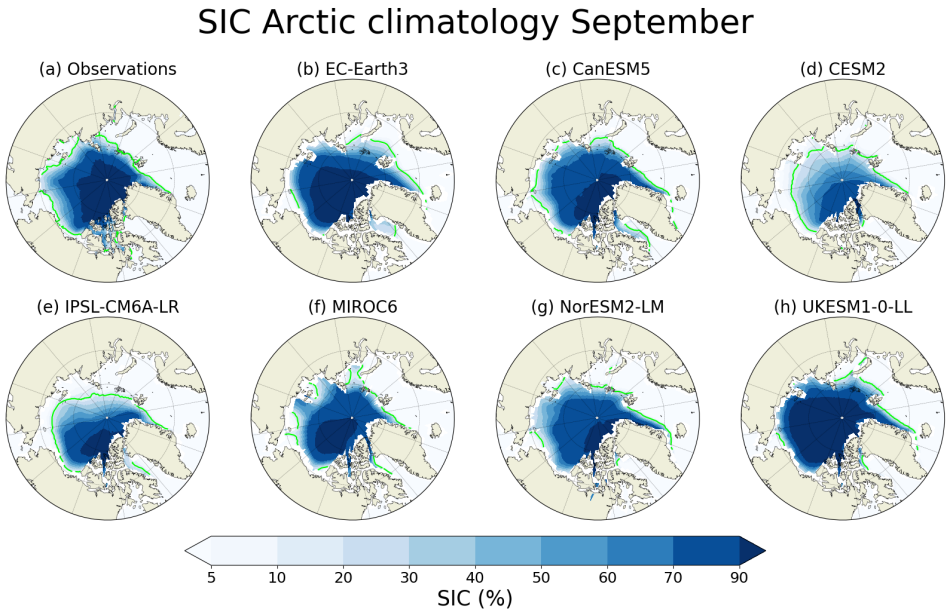

FIG. 5: Arctic September SIC climatology of the (a) HadISST observations and the (b) CMIP6 models. The lime contour indicates the sea ice edge.

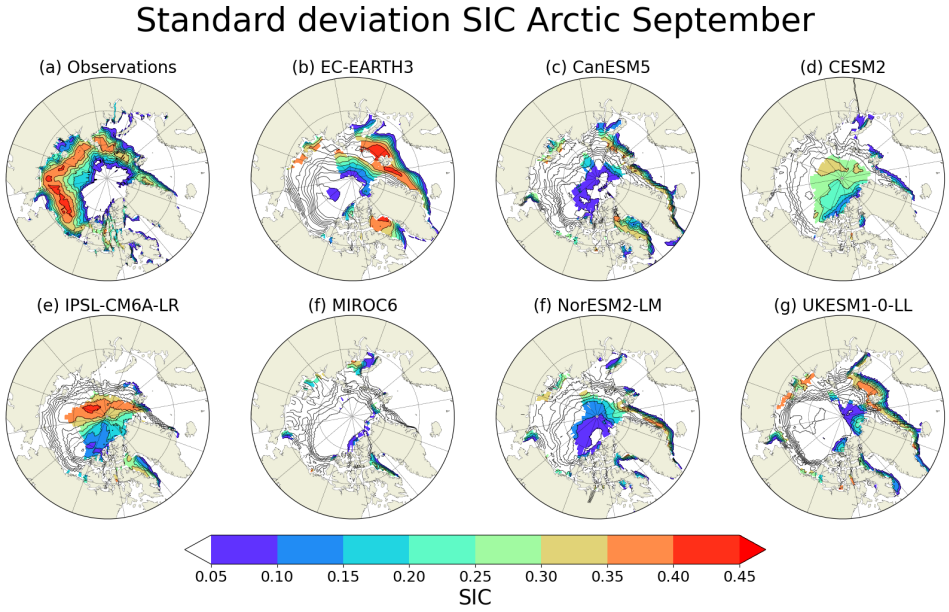

FIG. 6: Arctic September SIC standard deviation of the (a) HadISST observations and the (b) CMIP6 models. Black contours represent the total standard deviation, and only statistically significant standard deviation, using an f-test test at a 95% confidence level, is shaded.