Figures and Tables for the Revision of – *Detection and Tracking of Carbon Biomes via Integrated Machine Learning*

Sweety Mohanty[1,2], Lavinia Patara[1], Daniyal Kazempour[2], and Peer Kröger[2]

[1]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
[2]University of Kiel, Germany

Abstract.

Figures and Tables for the Revision

We are displaying the figures in the order they appear in the revised manuscript. Below we list which changes have been made to the figures.

5
- – The Fig 1, 2, 3, A1 and the Table B1 are unchanged.

- – We have updated Fig 5, 6, and 7 with correctly tracked biomes for 2010.

- – We have added Fig 8 showing the change in percentage coverage of carbon biomes between 1970-1979 and 2009-2018.

- – We have created a new Table E1 in appendix E to list correlations and linear trends of percentage coverage of selected biomes in the North Atlantic Ocean (NATL) and in the Southern Ocean (SO).

10
- – We have added Fig F1 in appendix F showing the change in SST between 1970-1970 and 2009-2018.

- – We have added Fig G1 in appendix G showing the comparison of simulated and observed climatologies of sea surface temperature, salinity, winter mixed layer depth, surface DIC and surface alkalinity.

- – We have added Fig H1 in appendix H showing the distribution of normalized regression coefficients (RCs) resulting from the MVLR between fCO_2 and its drivers.
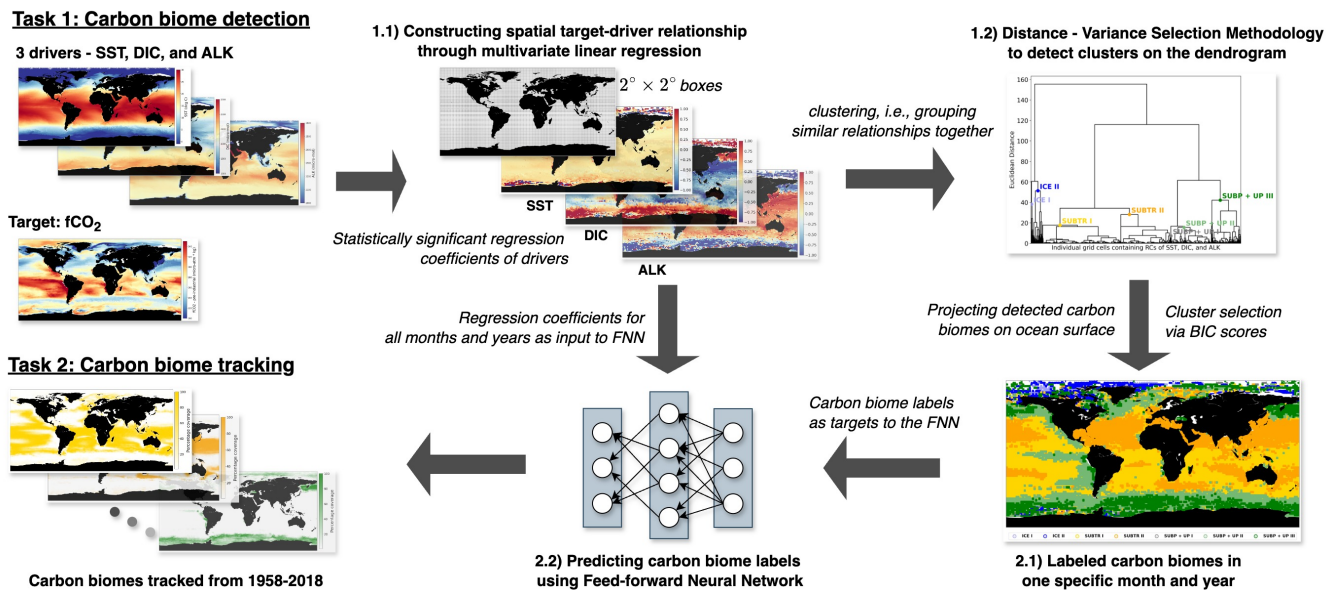
Figure 1. Schematic presentation of the step-by-step approach taken by our study to detect (Task 1) and track (Task 2) marine carbon biomes.
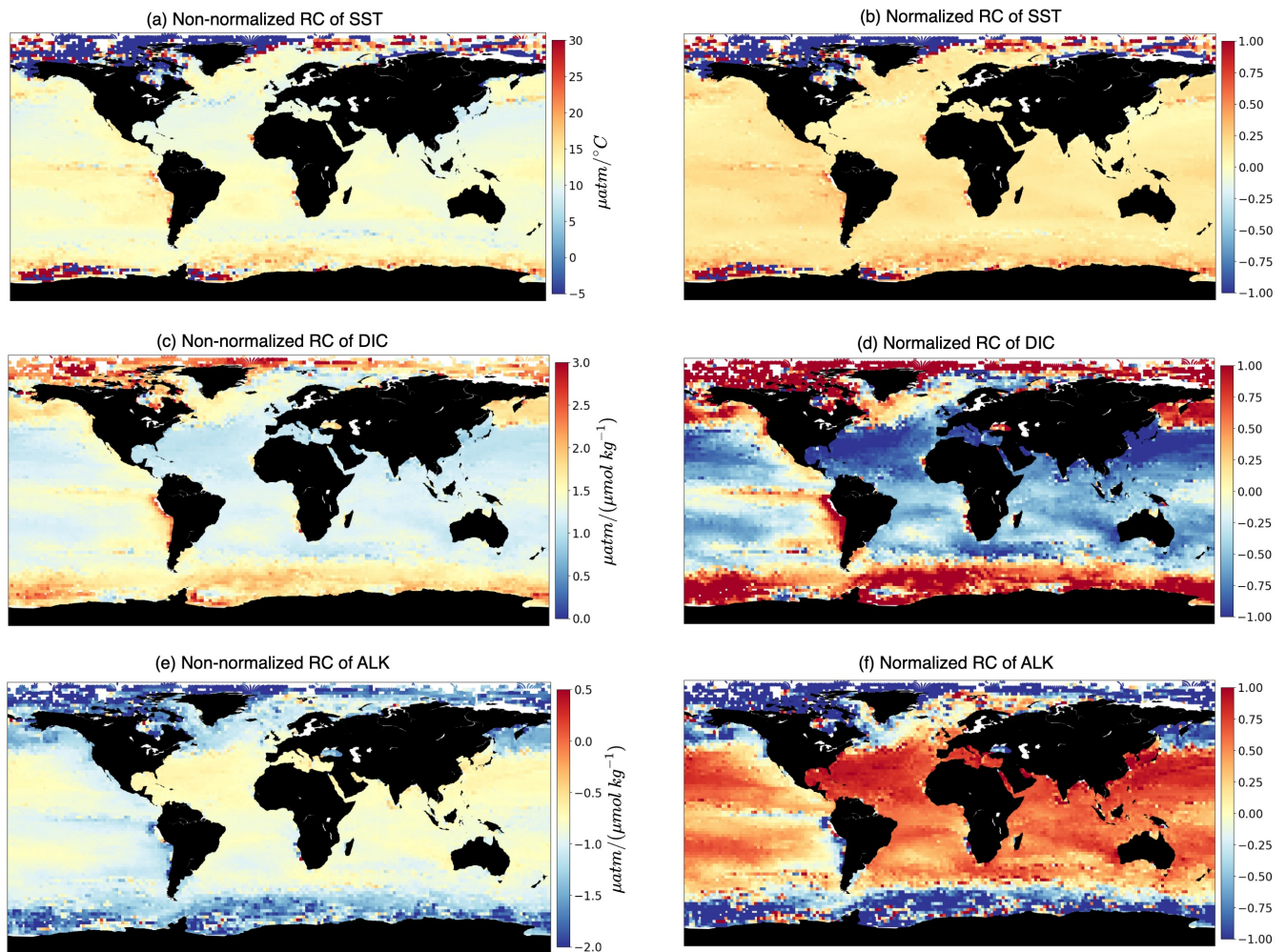
Figure 2. Spatial multivariate linear regression coefficients. The maps show non-normalized (left column) and normalized (right column) regression coefficients (RCs) computed over $2° \times 2°$ boxes of fCO_2 with respect to its three drivers, i.e. (a,b) SST, (c,d) surface DIC, and (e,f) surface ALK.
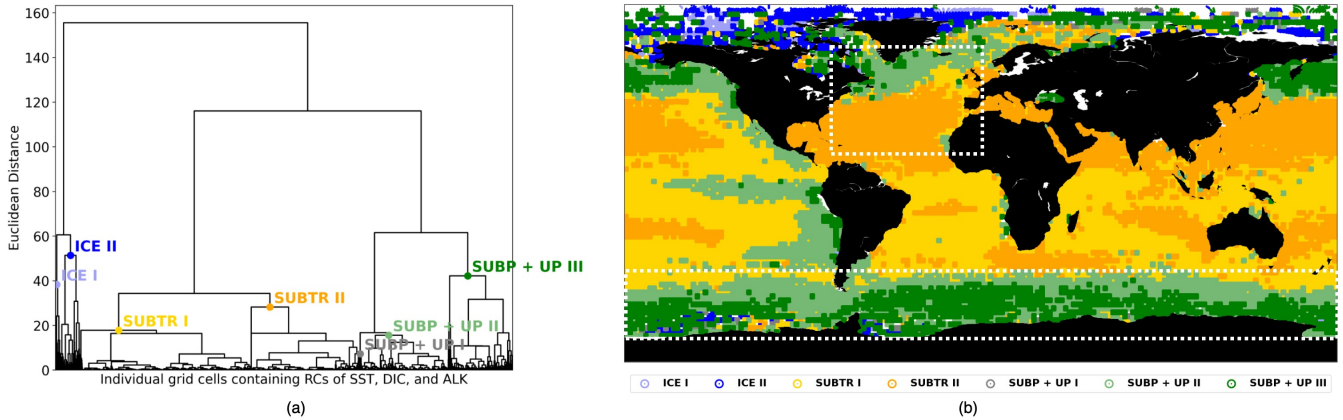
Figure 3. Carbon biomes in January 2009 detected through Hierarchical Clustering (HC) (a) Dendrogram resulting from the HC, with local cuts based on the distance-variance selection methodology. The text indicates the names of the detected seven clusters (i.e., the carbon biomes). (b) Geographical location of the detected clusters. The white boxes illustrate the basins - the North Atlantic between 75°W-0°, and 10°N-70°N and the Southern Ocean between 45°S-77°S.
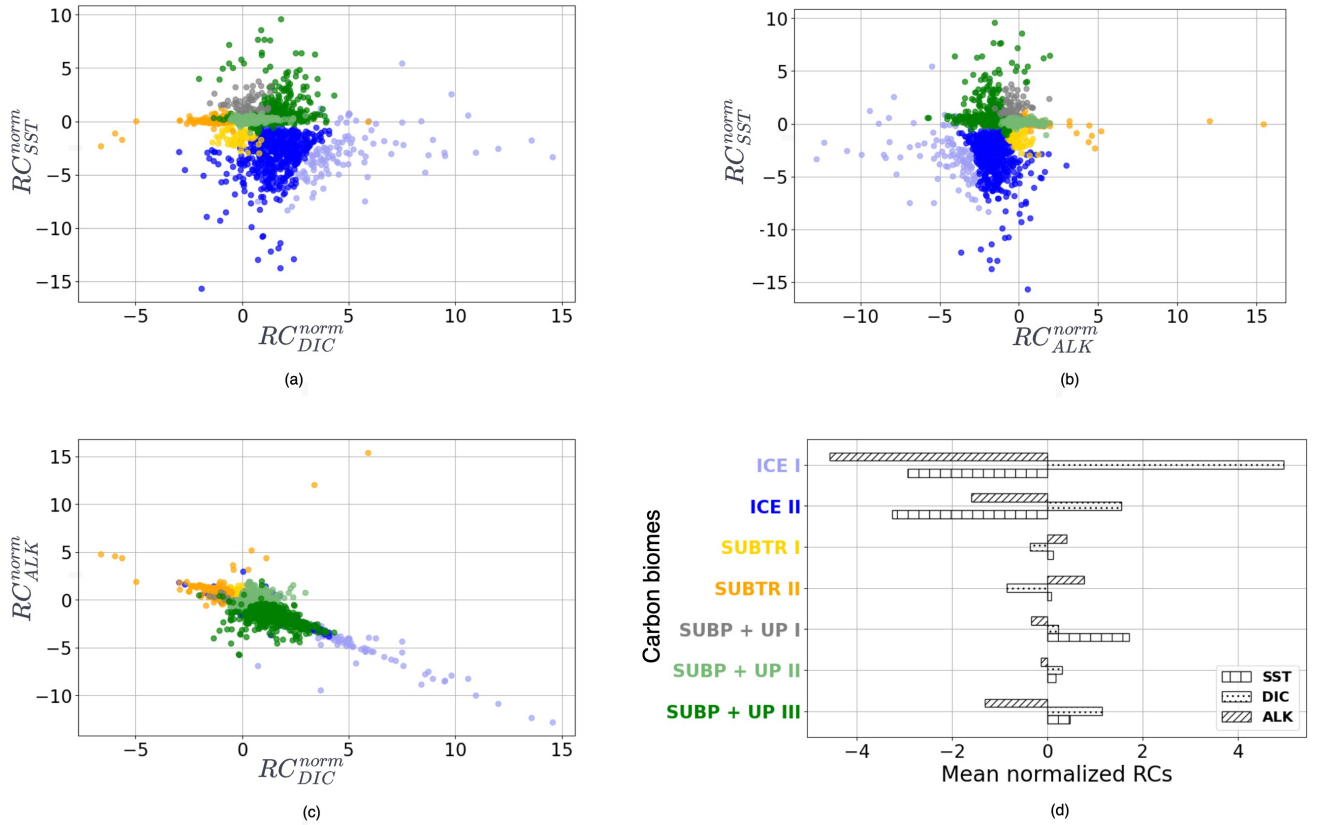
Figure 4. Distribution of normalized regression coefficients (RC^{norm}) in the carbon biomes detected for January 2009. (a) RC$_{SST}^{norm}$ against RC$_{DIC}^{norm}$, (b) RC$_{SST}^{norm}$ against RC$_{ALK}^{norm}$, (c) RC$_{ALK}^{norm}$ against RC$_{DIC}^{norm}$. (d) RC$_{SST}^{norm}$, RC$_{DIC}^{norm}$, and RC$_{ALK}^{norm}$ averaged over the carbon biomes. Colors indicate the detected carbon biomes: ICE biomes in blue shading, SUBP+UP biomes in green shading, and SUBTR biomes in orange shading. SUBP+UP I was not tracked (see main text) and is therefore shown in grey.

References

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, Journal of Geophysical Research: Oceans, 118, 6704–6716, 2013.

30 Lauvset, S. K., Key, R. M., Olsen, A., Van Heuven, S., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., et al.: A new global interior ocean mapped climatology: The 1×1 GLODAP version 2, Earth System Science Data, 8, 325–340, 2016.

Sallée, J.-B., Pellichero, V., Akhoudas, C., Pauthenet, E., Vignes, L., Schmidtko, S., Garabato, A. N., Sutherland, P., and Kuusela, M.: Summertime increases in upper-ocean stratification and mixed-layer depth, Nature, 591, 592–598, 2021.
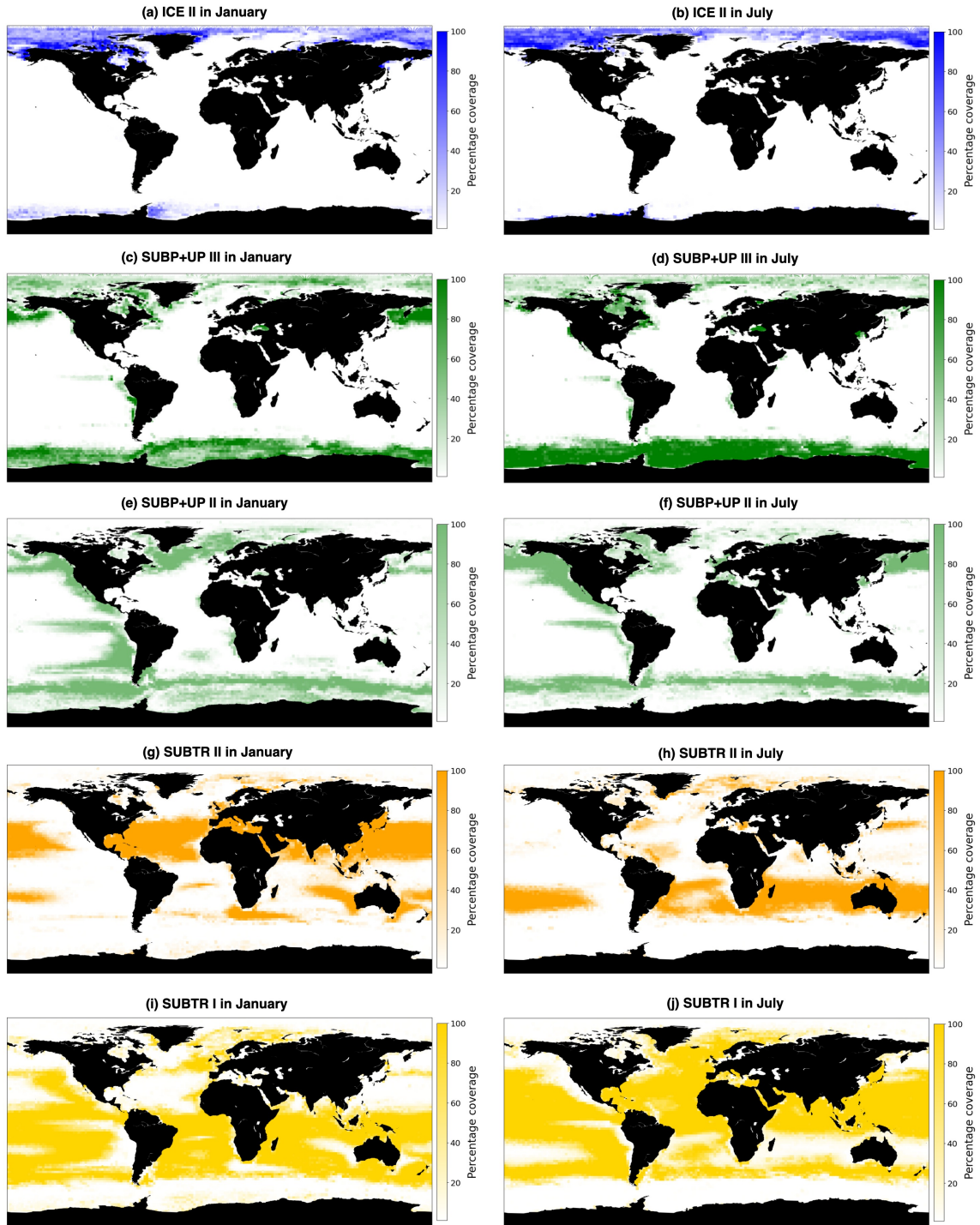
Figure 5. Percentage coverage of carbon biomes over the years 1958 to 2018, where a value of 100% indicates that the biome is present in each simulation year in the 2x2 box and a value of 0% that it is never present in the 2x2 box. Shown are the main 5 carbon biomes: a,b) ICE II, c,d) SUBP+UP III, e,f) SUBP+UP II, g,h) SUBTR II, and i,j) SUBTR I in January (left column) and July (right column).
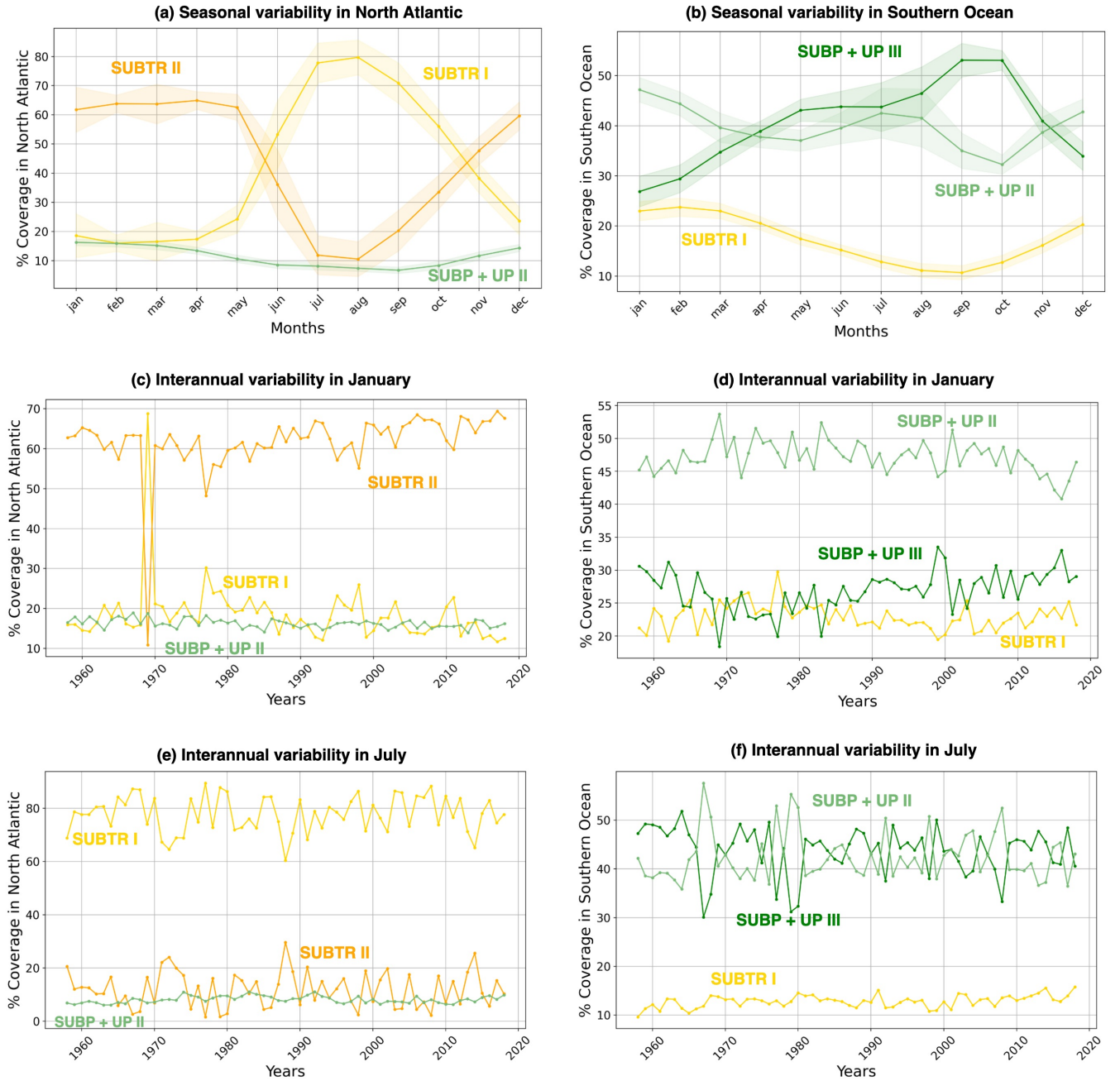
Figure 6. Seasonal and interannual variations of carbon biomes coverage. Shown is the percentage coverage of the three main biomes over the North Atlantic (left column) and Southern Ocean (right column). The limits of the basins are shown in Fig. 3 b. The percentage coverage is computed as the weighted area average of the carbon biome divided by the total area of the basin so that a value of 100% would be achieved when the biome covers the whole basin. (a,b) Mean seasonal cycle of the carbon biome coverage, with shading indicating the standard deviation across the 61 simulation years. (c-f) Weighted area average of the percentage coverage of carbon biomes over 1958-2018 in the months of c,d) January and e,f) July. Here, the North Atlantic basin lies between 75°W-0°, and 10°N-70°N and the Southern Ocean is between 45°S-77°S.
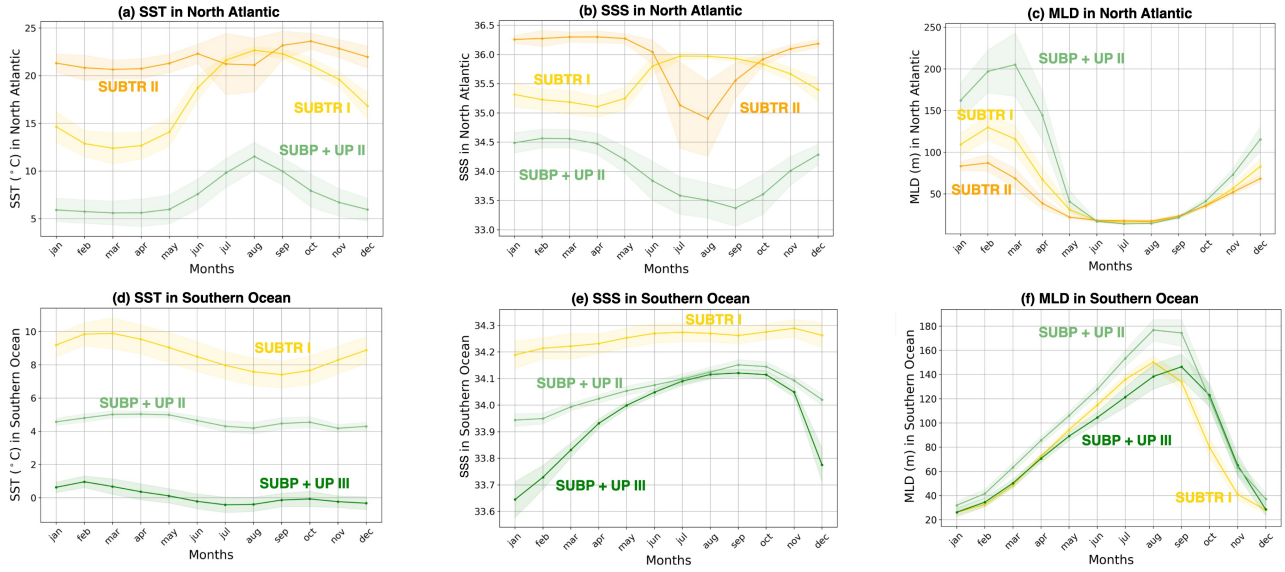
7

Figure 7. Seasonality of environmental parameters over biomes in the North Atlantic and Southern Ocean. Mean seasonal cycle over 1958-2018 of (a,d) SST, (b,e) sea surface salinity (SSS), and (c,f) mixed layer depth (MLD), averaged within the three main carbon biomes in the North Atlantic (top row) and Southern Ocean (bottom row). The limits of the basins are shown in Fig. 3 b. Shading indicates the standard deviation over the 61 simulation years. The North Atlantic basin lies between $75°$W-$0°$, and $10°$N-$70°$N and the Southern Ocean is between $45°$S-$77°$S.
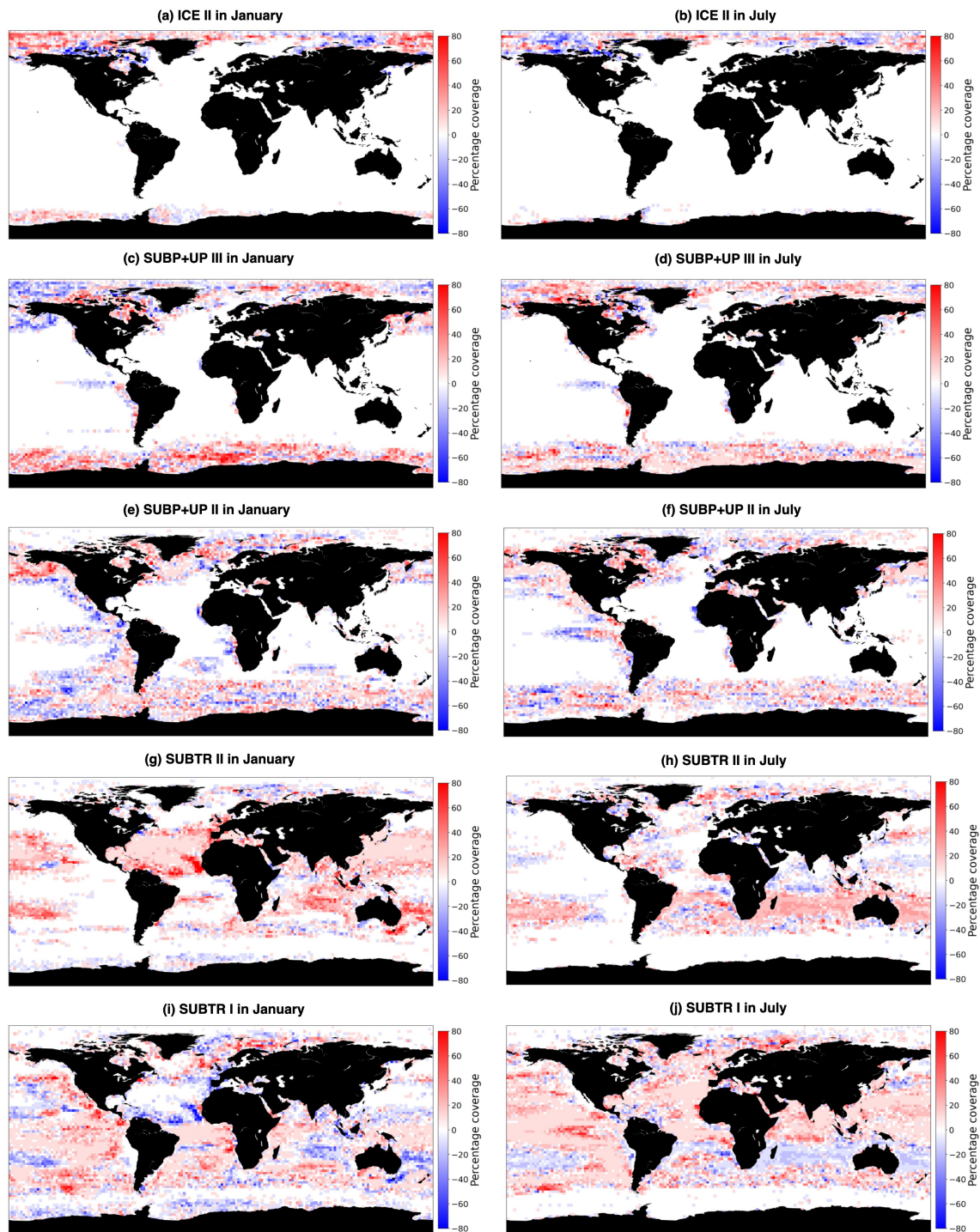
Figure 8. Change in percentage coverage of carbon biomes between 1970-1979 and 2009-2018. Shown are the main 5 carbon biomes: a,b) ICE II, c,d) SUBP+UP III, e,f) SUBP+UP II, g,h) SUBTR II, and i,j) SUBTR I in January (left column) and July (right column). As an example, a value of +100% indicates that over a 2° box, a particular biome was present for each of the 10 years from 2009-2018 and never present in each of the 10 years between 1970-1979. Conversely, a value of -100% indicates that the biome was always present from 1970-1979 and never in 2009-2018. A value of 0% indicates that the biome was equally present between the two decades.
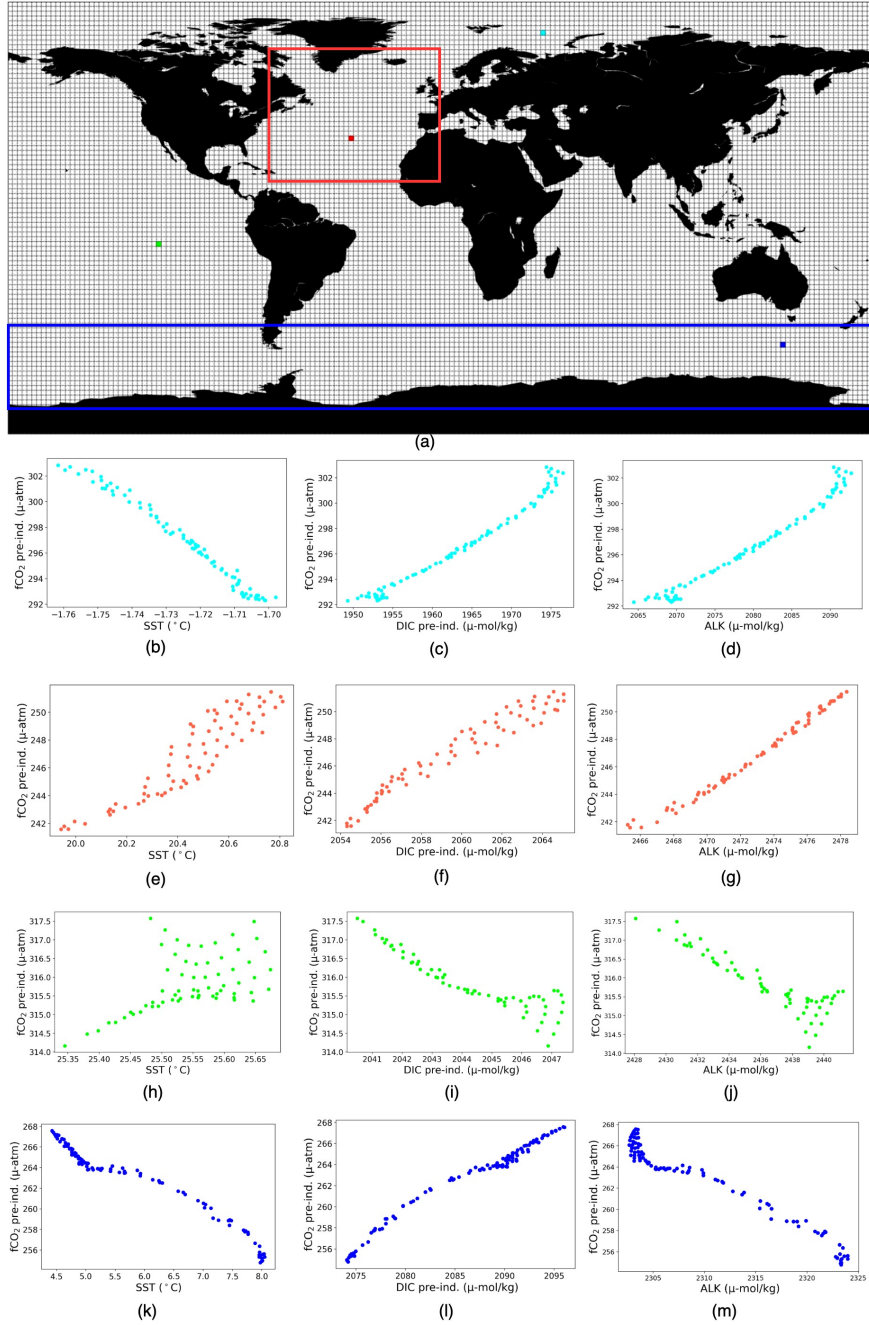
Figure A1. (a) $2° \times 2°$ boxes used for computing multivariate linear regressions, with the contours indicating the two basins - the North Atlantic and Southern Ocean. b-m: scatter plots of fCO$_2$ against SST (first column), DIC (second column) and ALK (third column) over the four boxes indicated as colors in panel a.

Sr. No.	Month	NATL			SO		
		Correlations (SUBTR I vs. II)	Trend SUBTR I ($\% \, decade^{-1}$)	Trend SUBTR II ($\% \, decade^{-1}$)	Correlations (SUBP+UP II vs. III)	Trend SUBP+UP II ($\% \, decade^{-1}$)	Trend SUBP+UP III ($\% \, decade^{-1}$)
1	Jan	**-0.96**	**-1.44**	**1.87**	**-0.65**	**-0.79**	**1.34**
2	Feb	**-0.96**	**-0.95**	**1.31**	**-0.62**	**-0.63**	**1.0**
3	Mar	**-0.90**	**-0.87**	**1.09**	**-0.74**	**-0.45**	**0.73**
4	Apr	**-0.77**	**-0.79**	**1.13**	**-0.50**	-0.07	0.44
5	May	**-0.59**	-0.33	0.64	**-0.63**	0.18	0.08
6	Jun	**-0.81**	0.91	-0.31	**-0.68**	0.04	-0.09
7	Jul	**-0.65**	0.78	-0.4	**-0.75**	-0.17	-0.03
8	Aug	**-0.60**	-0.27	0.36	**-0.76**	-0.08	-0.2
9	Sep	**-0.60**	0.52	-0.4	**-0.61**	-0.11	-0.05
10	Oct	**-0.40**	0.41	-0.09	**-0.46**	0.24	-0.1
11	Nov	**-0.74**	0.18	0.47	**-0.57**	-0.11	0.41
12	Dec	**-0.87**	**-1.08**	**1.59**	**-0.48**	**-0.68**	**0.88**

Table E1. Correlations and linear trends of percentage coverage of selected biomes in North Atlantic Ocean (NATL) and in the Southern Ocean (SO). The percentage coverage is computed as percentage area covered by a specific biome over each basin. For each month over the period 1970-2018, correlations coefficients are computed between the percentage coverage of SUBTR I and SUBTR II in NATL and between the percentage coverage of SUBP+UP II and SUBP+UP III in the SO. The linear trends of biome coverage of SUBTR I and SUBTR II (for NATL) and of SUBP+UP II and SUBP+UP III (for SO) are computed over the period 1970-2018 separately for each month. For both correlation coefficients and linear trends, statistically significant values with p-value less than 0.04 are highlighted in bold
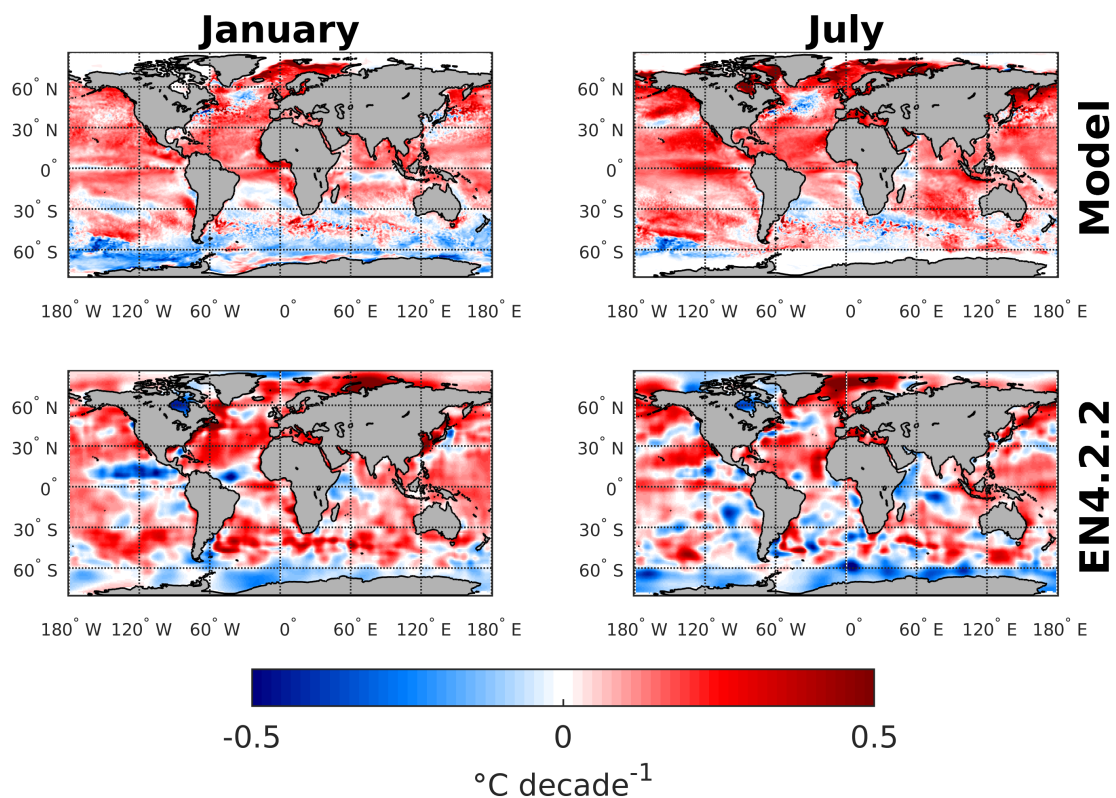
.

Figure F1. Change in SST between 1970-1970 and 2009-2018 in the model and in the EN4.2.2 observation-based product (**?**). Top row: model, bottom row: EN4.2.2 product; left column: January, right column: July. The trend value in $^\circ$C decade^{-1} has been computed by dividing the SST difference by the number of years separating the median years of the two decades (i.e. 39 years) and multiplying by 10.
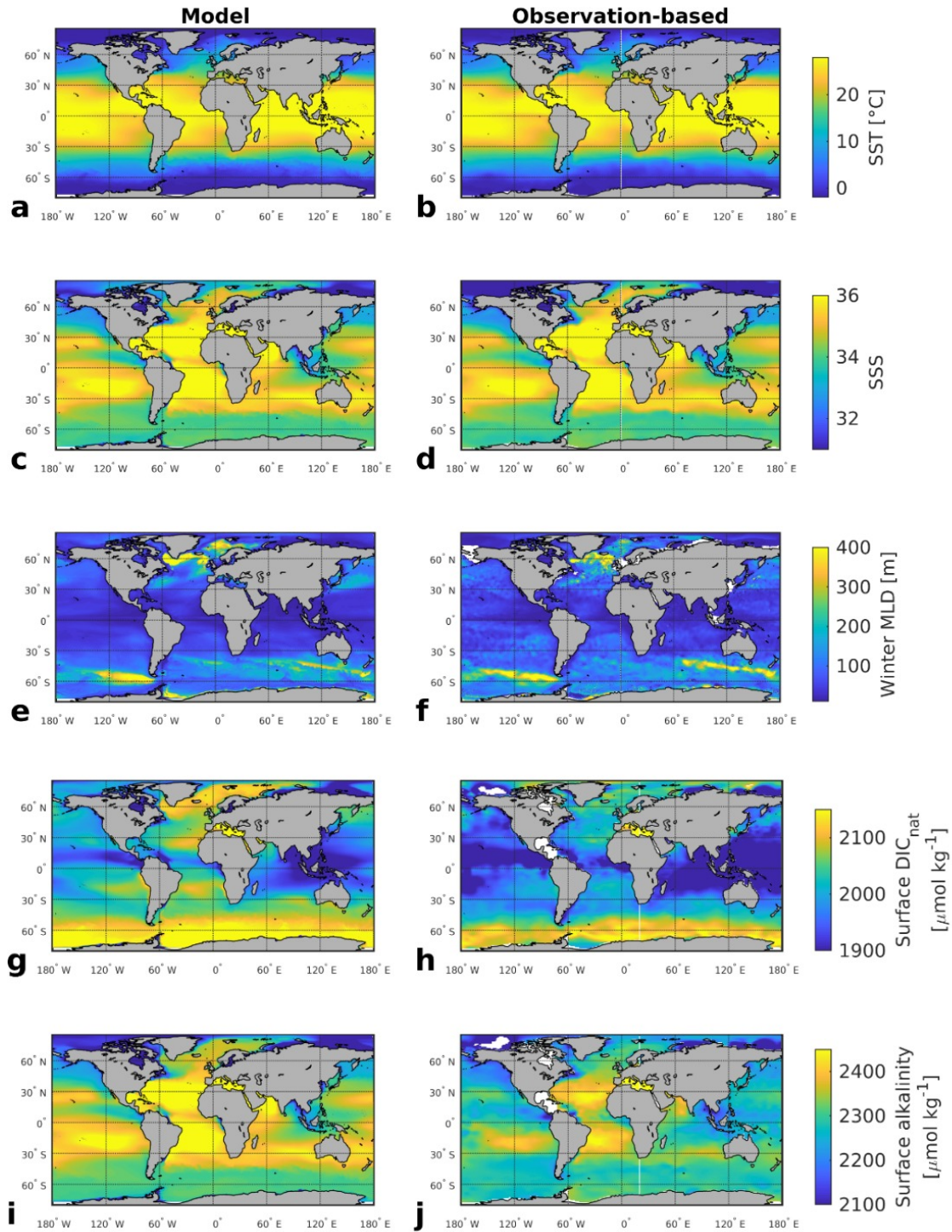
Figure G1. Comparison of simulated and observed climatologies of a,b) Sea surface temperature (SST), c,d) Sea surface salinity (SSS), e,f) Mixed Layer Depth (MLD) in March for the Northern Hemisphere and in September for the Southern Hemisphere, g,h) natural (or pre-industrial) surface dissolved inorganic carbon (DIC), i,j) surface alkalinity. Observational estimates are Good et al. (2013) for SST and SSS, Sallée et al. (2021) for MLD, GLODAPv2 2016 release (Lauvset et al., 2016) for pre-industrial DIC and Alkalinity. Shown are 2000-2018 climatologies for all variables except MLD, for which a 1970-2018 average was computed (in accordance with the observational data set).
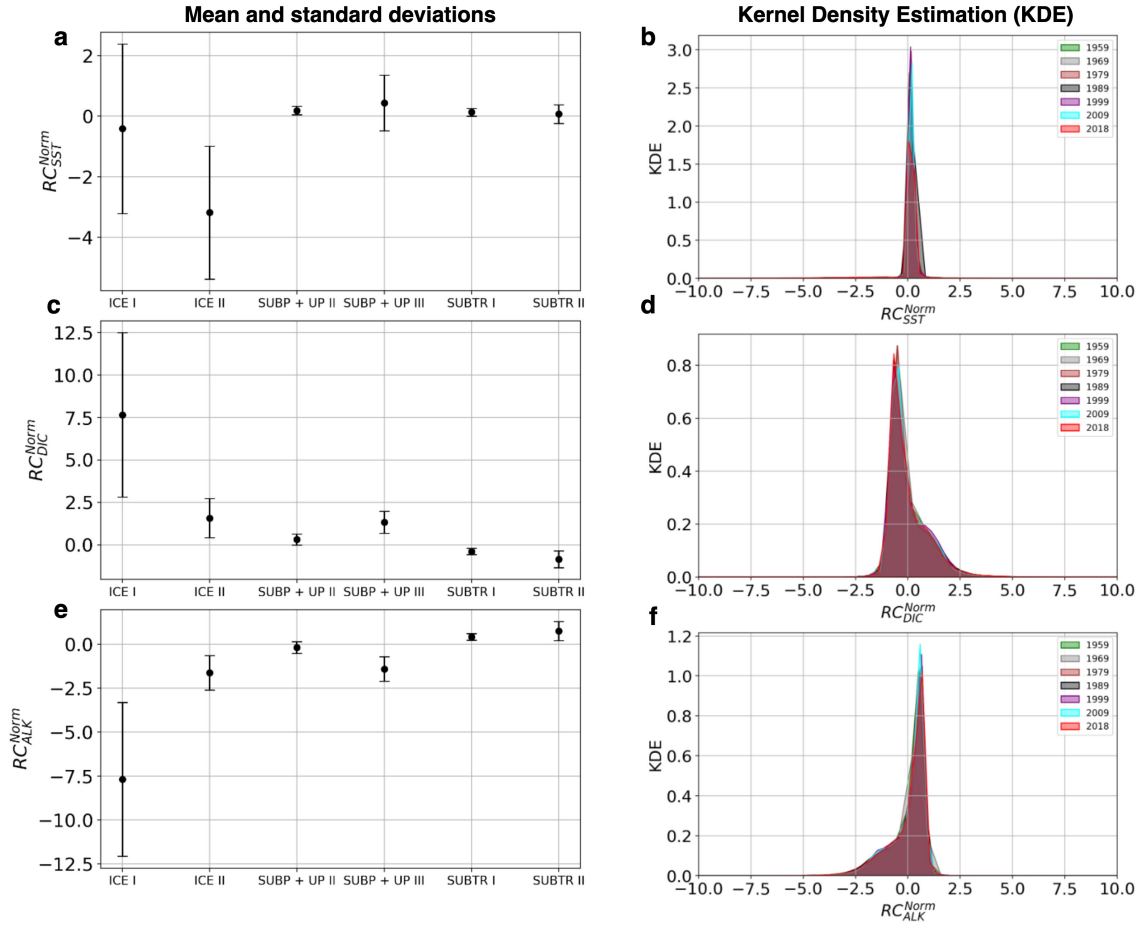
Figure H1. Distribution of normalized regression coefficients (RCs) resulting from the MVLR between fCO$_2$ and its drivers, i.e. a,b) SST, c,d) DIC, and e,f) ALK. Left column: mean RCs over the each of the 6 tracked biomes ICE I, ICE II, SUBP+UP II, SUBP+UP III, SUBTR I and SUBTR II. Shown is the mean and the standard deviation computed over each of the 12 months between 1958 and 2018. Right column: Kernel density estimation (KDE) of normalized RCs over selected years. The KDE plot estimates the probability density function (PDF) of a continuous random variable, in our case, the RCs, and visualizes the distribution of data points in a smooth curve, which helps to understand the underlying data distribution.