Supplement

Assessing improvements in global ocean pCO_2 machine learning reconstructions with Southern Ocean autonomous sampling

Thea H. Heimdal[1], Galen A. McKinley[1], Adrienne J. Sutton[2], Amanda R. Fay[1], Lucas Gloege[3]
[1]Columbia University and Lamont-Doherty Earth Observatory, Palisades, NY, USA
[2]Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, WA, USA
[3]Open Earth Foundation, Marina del Rey, CA, USA

Correspondence to: Thea H. Heimdal (theimdal@ldeo.columbia.edu)

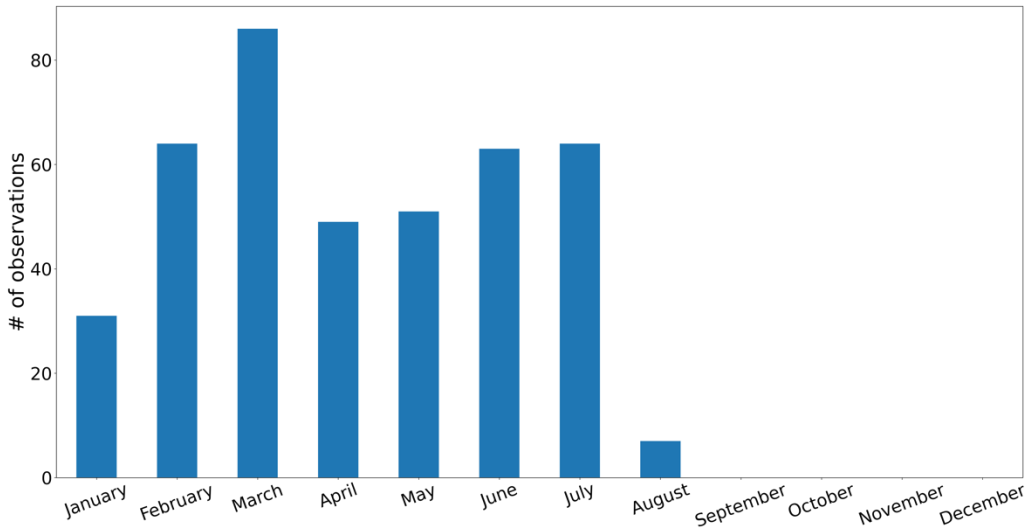

Figure S1: Number of monthly 1°x1° observations from the 'real-world' Saildrone USV journey in Sutton et al. (2021).

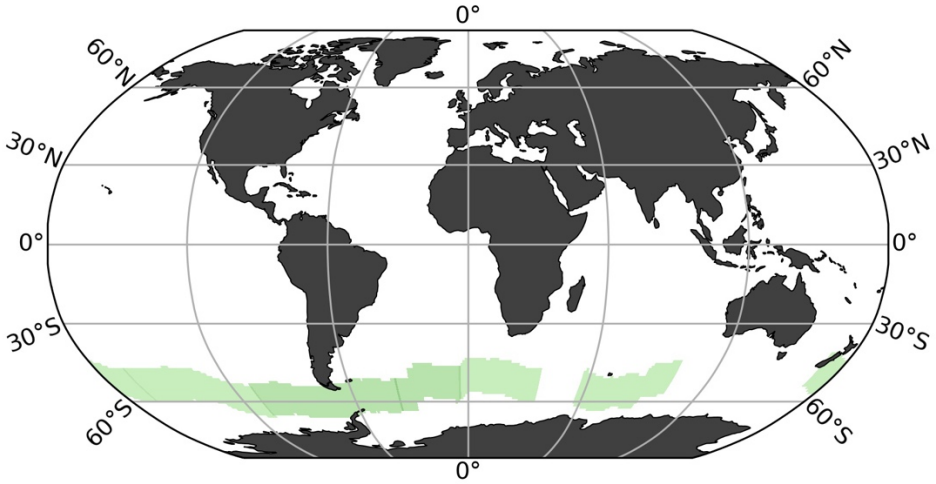

Figure S2: Map showing the spatial extent of sampling with 13 Saildrone USVs for run 'x13_10Y_J-A' and 'x13_10Y_W'. The 'real-world' USV track from Sutton et al. (2021) was repeated six times by 1°, covering an additional 6° both north and south.

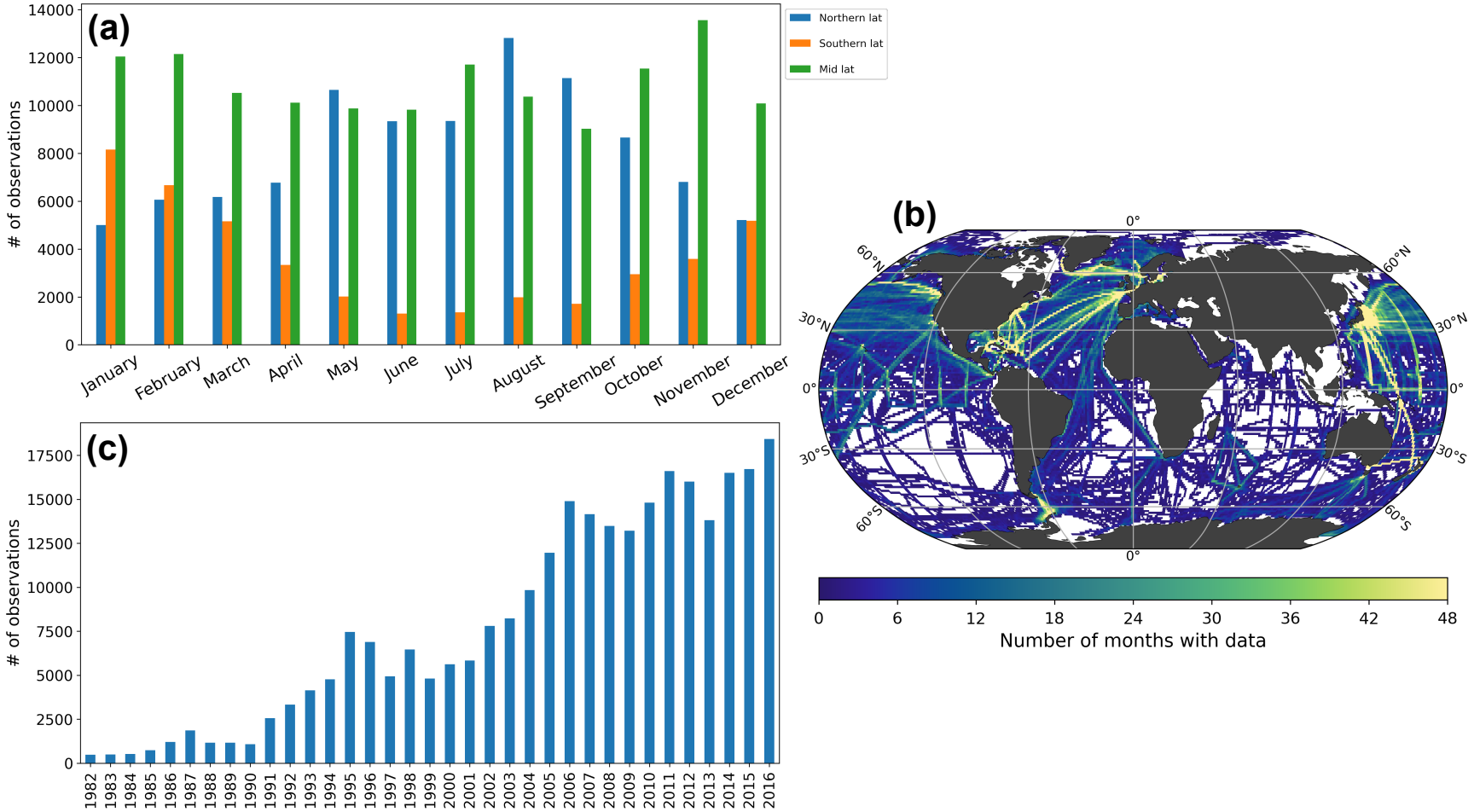

Figure S3: SOCAT observations from 1982-2016. Number of monthly 1°x1 SOCAT observations per month (**a**). Spatial extent of SOCAT tracks and number of monthly observations per 1°x1° grid cell (**b**). Number of monthly 1°x1 SOCAT observations per year (**c**).

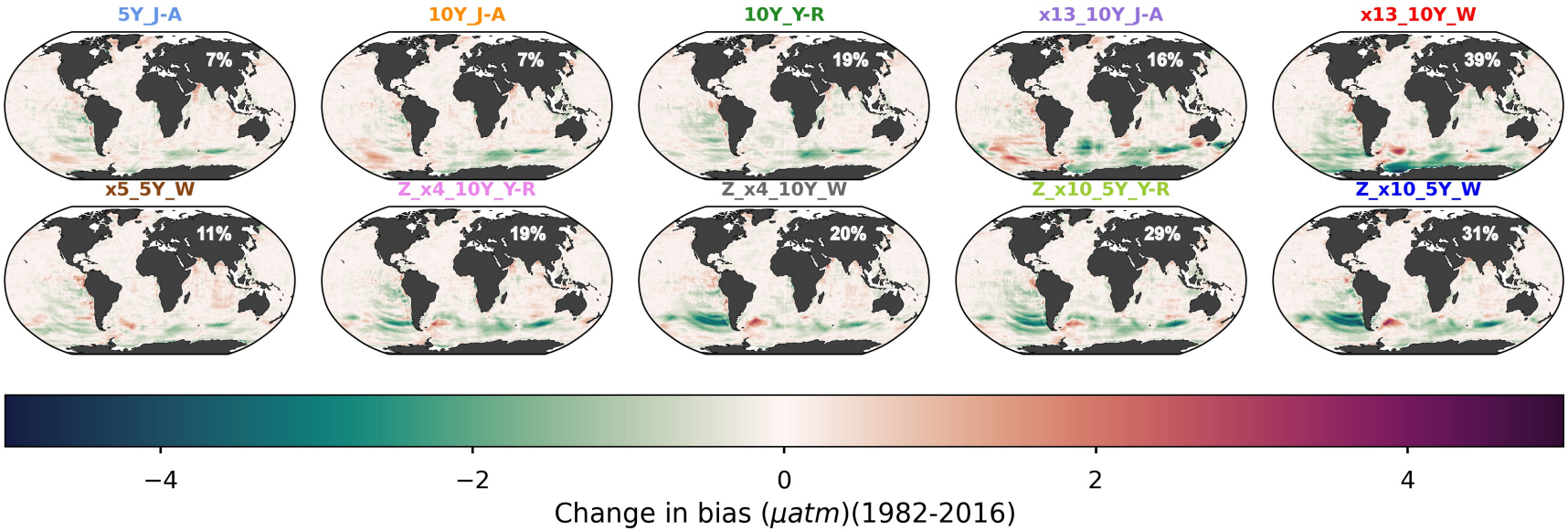

Figure S4. Change in bias when comparing the different Saildrone USV sampling strategies (see **Table 1**) to the baseline SOCAT reconstruction. The bias was calculated by comparing each machine learning reconstruction to the testbed 'model truth' averaged over the 75 ensemble members over the period of 1982 through 2016. Red and green areas represent regions where the change in bias is positive or negative, respectively. Generally, there is a negative change in bias in southern latitudes, indicating an improvement over the baseline that overestimated pCO_2 (**Fig. 3a**). The green shadings indicate therefore improvement in bias compared to using SOCAT data alone. The improvement compared to the SOCAT baseline is shown by percent values (global average for 1982-2016). The 'one-latitude'-run 'x13_10Y_W' demonstrates the most significant improvement (i.e., 39%).

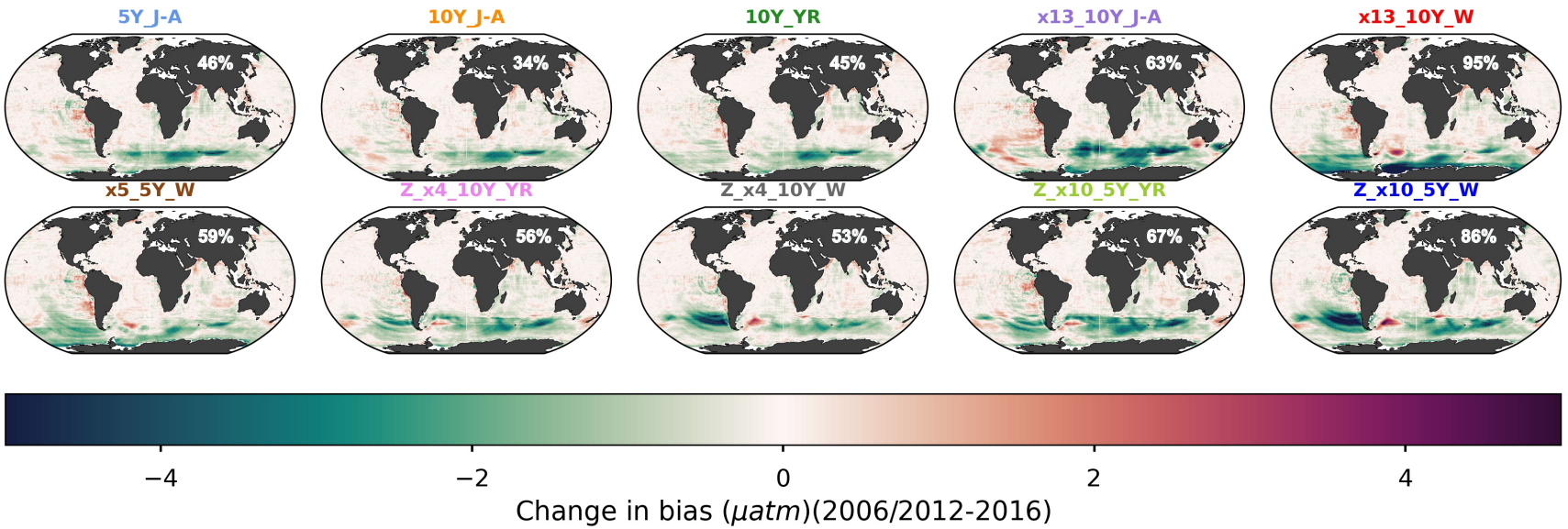

Figure S5: Same as **Figure A4**, but averaged over the years of USV sample addition (2006-2016 or 2012-2016). There is significant improvement in bias for all runs compared to the entire testbed period (1982-2016; **Fig. A4**).

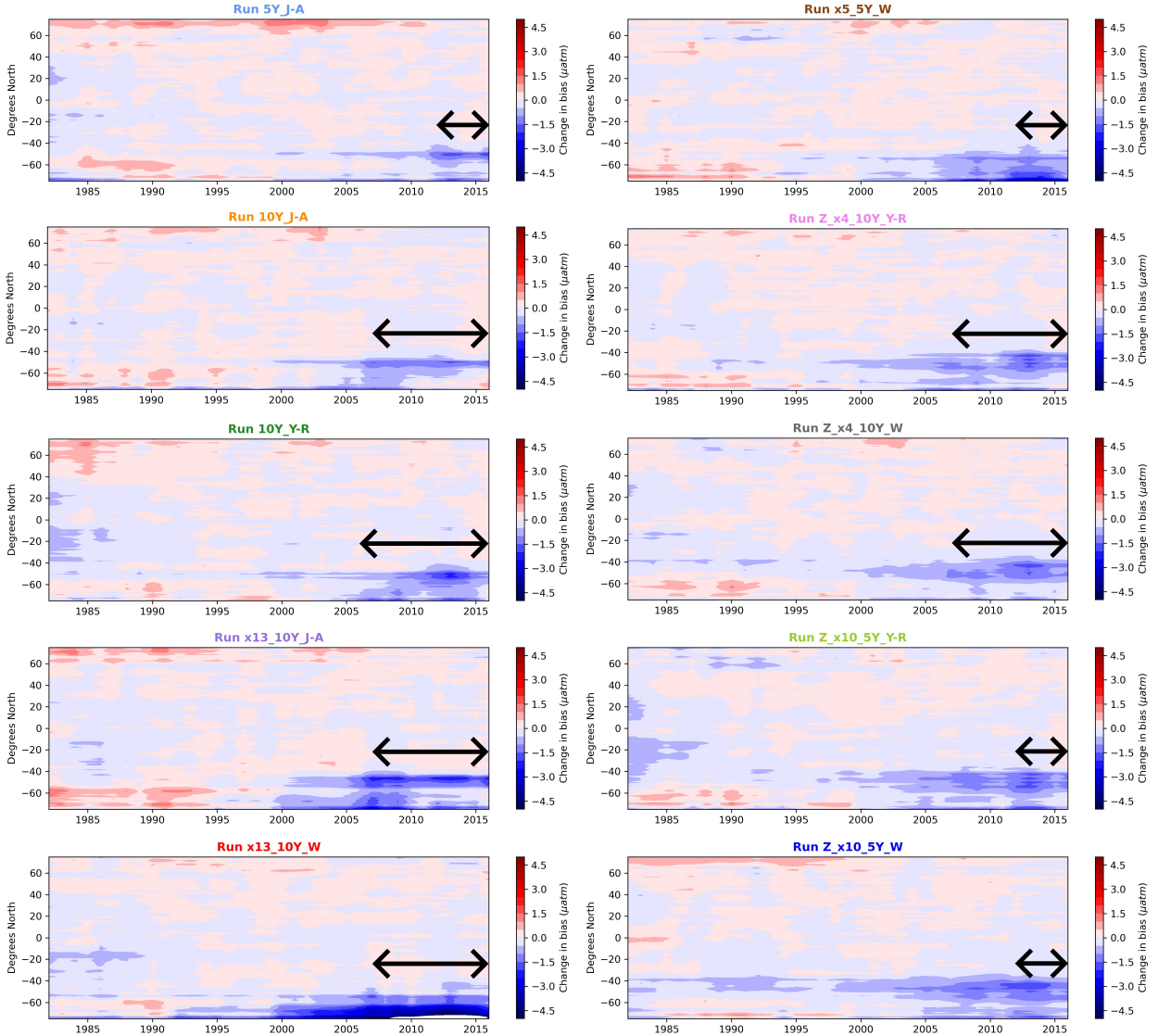

Figure S6: Zonal mean, annual mean Hovmöller of the change in bias when comparing the different Saildrone USV sampling strategies to the baseline SOCAT reconstruction. Red and blue areas represent regions where the change in bias is positive or negative, respectively. The negative change in bias in southern latitudes represents an improvement considering the overestimation of pCO_2 shown by the baseline reconstruction (**Fig. 6a**). Bias improvements expand backwards to year 2000 for all runs, which is well beyond the duration of USV additions (shown by arrow on panel).

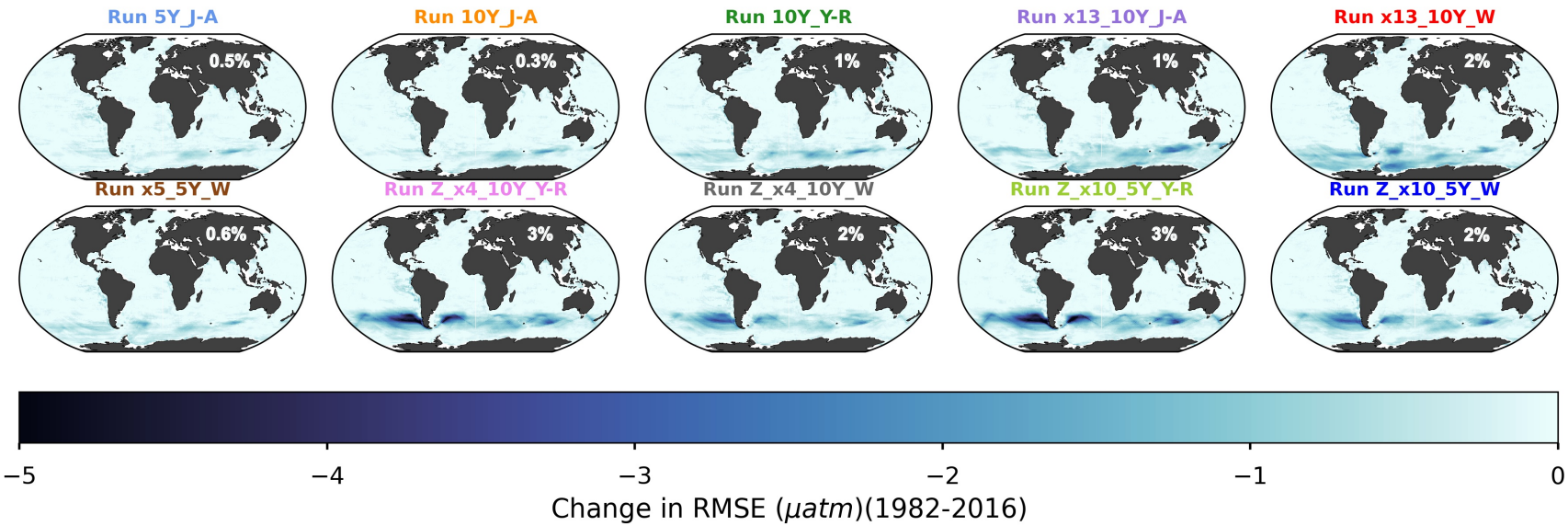

Figure S7: Change in RMSE when comparing the different Saildrone USV sampling strategies to the baseline SOCAT reconstruction. The RMSE was calculated by comparing each machine learning reconstruction to the testbed 'model truth', averaged over the 75 ensemble members over the period of 1982 through 2016. Dark areas represent regions with improvement in RMSE (i.e., decreasing RMSE). Improvement in RMSE occurs mainly in southern latitudes (<35°S), where the baseline reconstruction demonstrated high RMSEs (**Fig. 3b**). Percent values represent improvement compared to the SOCAT baseline.

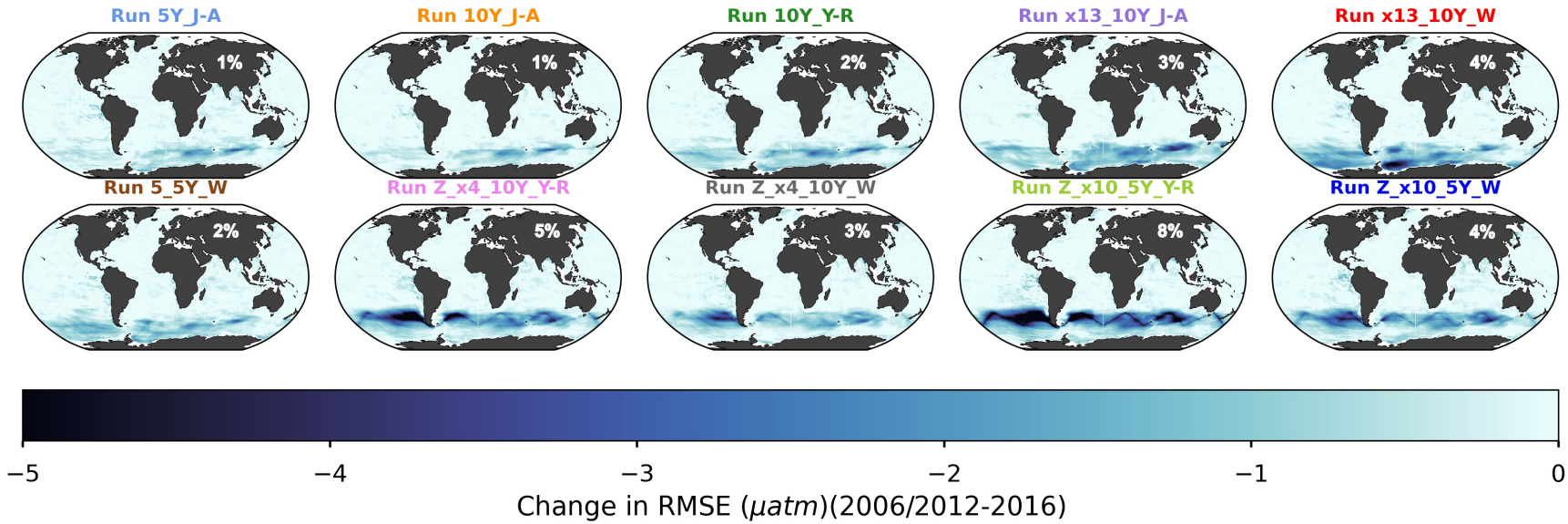

Figure S8: Same as **Figure A7**, but averaged over the years of USV sample addition (2006-2016 or 2012-2016). Compared to the mean of the entire testbed period (1982-2016; **Fig. A7**), there is improvement in bias for all runs. Percent values represent improvement compared to the SOCAT baseline.

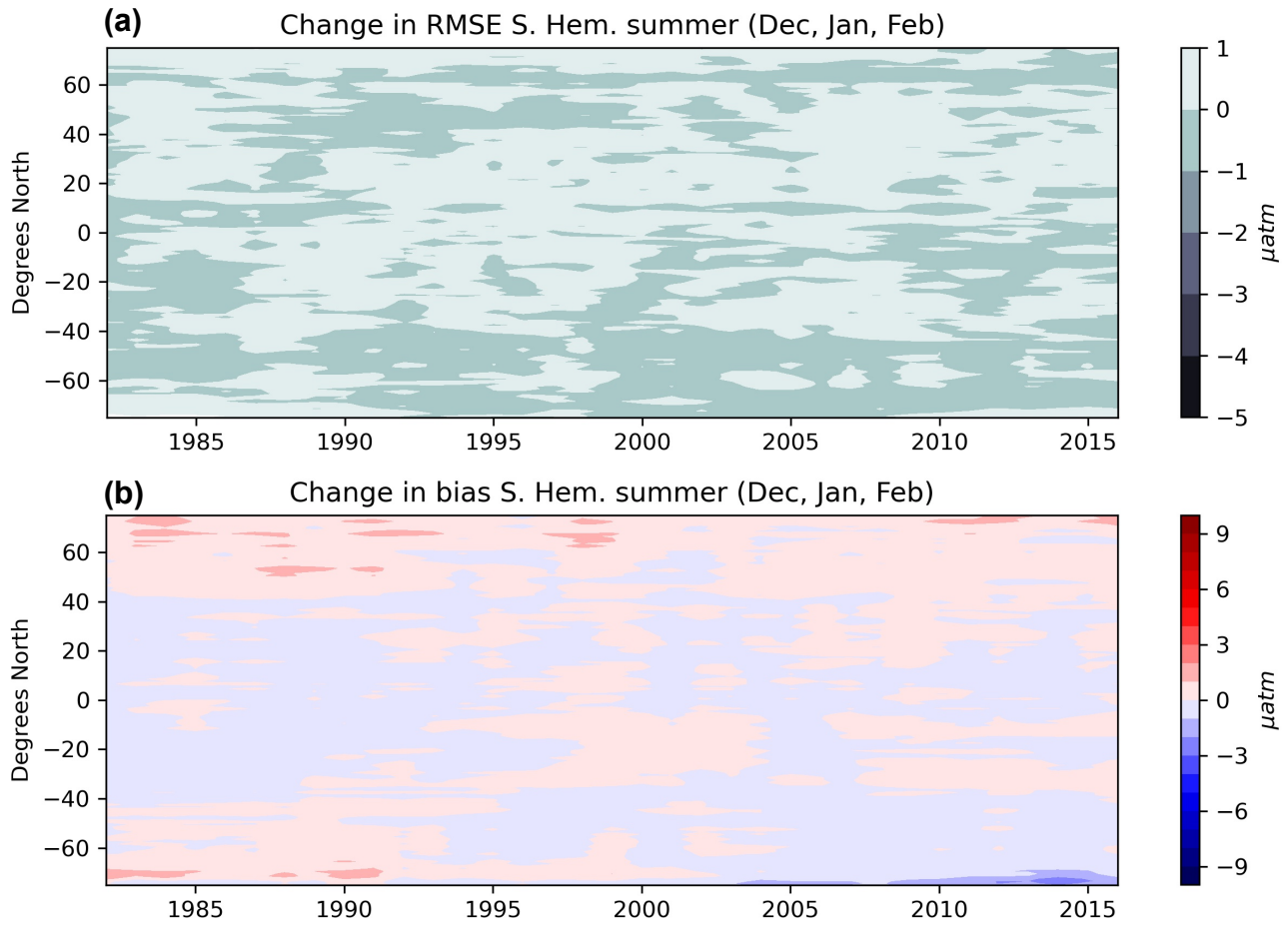

Figure S9: Zonal mean, DJF (December, January, February) mean Hovmöller change in RMSE (**a**) and bias (**b**) for run 'x13_10Y_W' compared to the baseline SOCAT reconstruction. There is minimal improvement in RMSE and bias during southern hemisphere summer months.

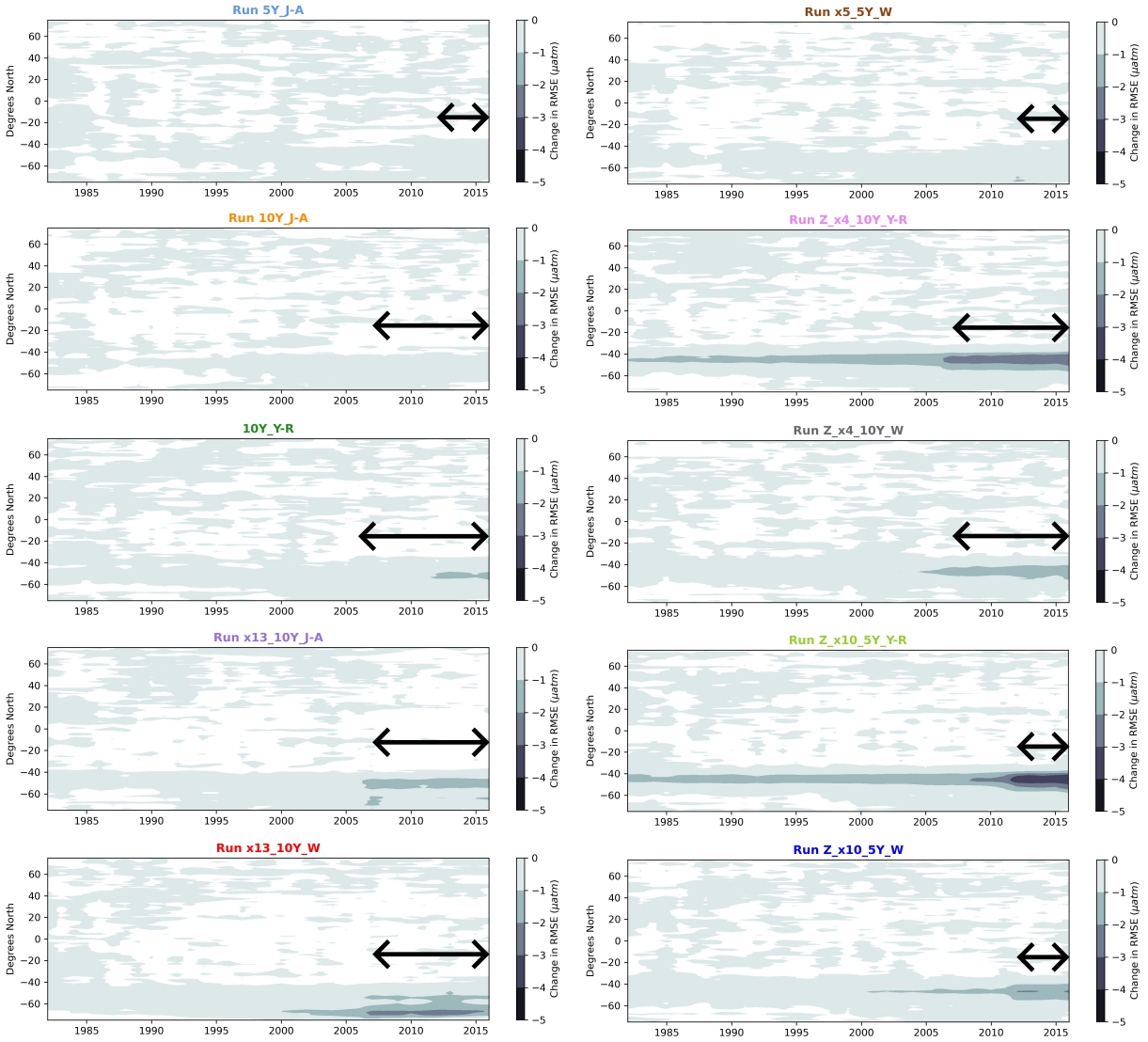

Figure S10: Zonal mean, annual mean Hovmöller of the change in RMSE when comparing the different Saildrone USV sampling strategies to the baseline SOCAT reconstruction. The most significant improvement in RMSE is shown for the 'zigzag'-runs and the 'one-latitude'-run 'x13_10Y_W' in the Southern Ocean during the last 5-10 years when the additional Saildrone observations are introduced (shown by arrow on panel). For the 'year-round' 'zigzag' runs, improvement expands back in time to the beginning of the testbed period.

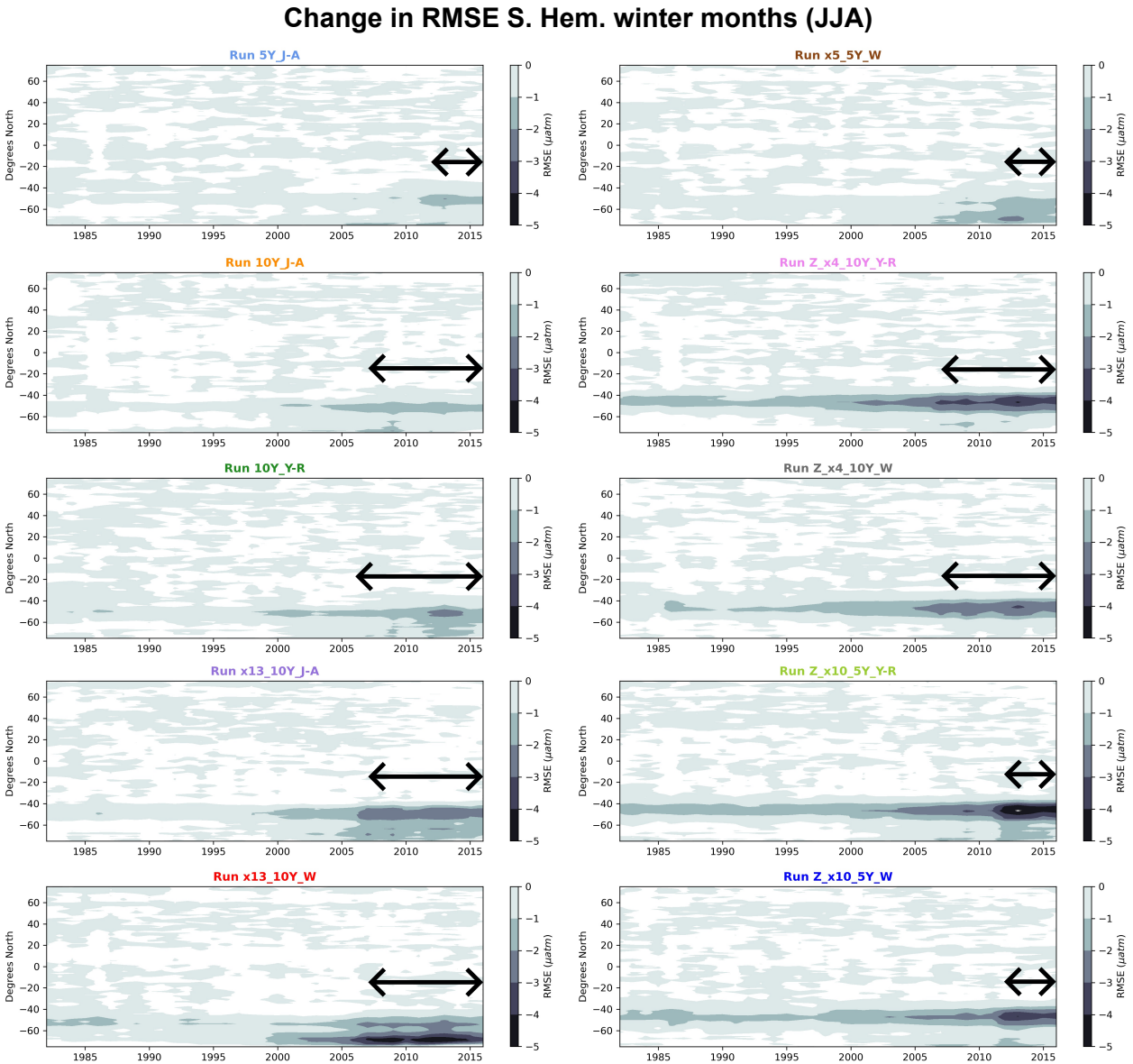

Figure S11: Same as **Figure A10**, but for southern hemisphere winter months (i.e., JJA; June, July, August). Compared to **Figure A10** (averaging over all months), there is improvement in RMSE both in terms of the magnitude and duration. Arrows represent the timing of Saildrone USV sample additions.

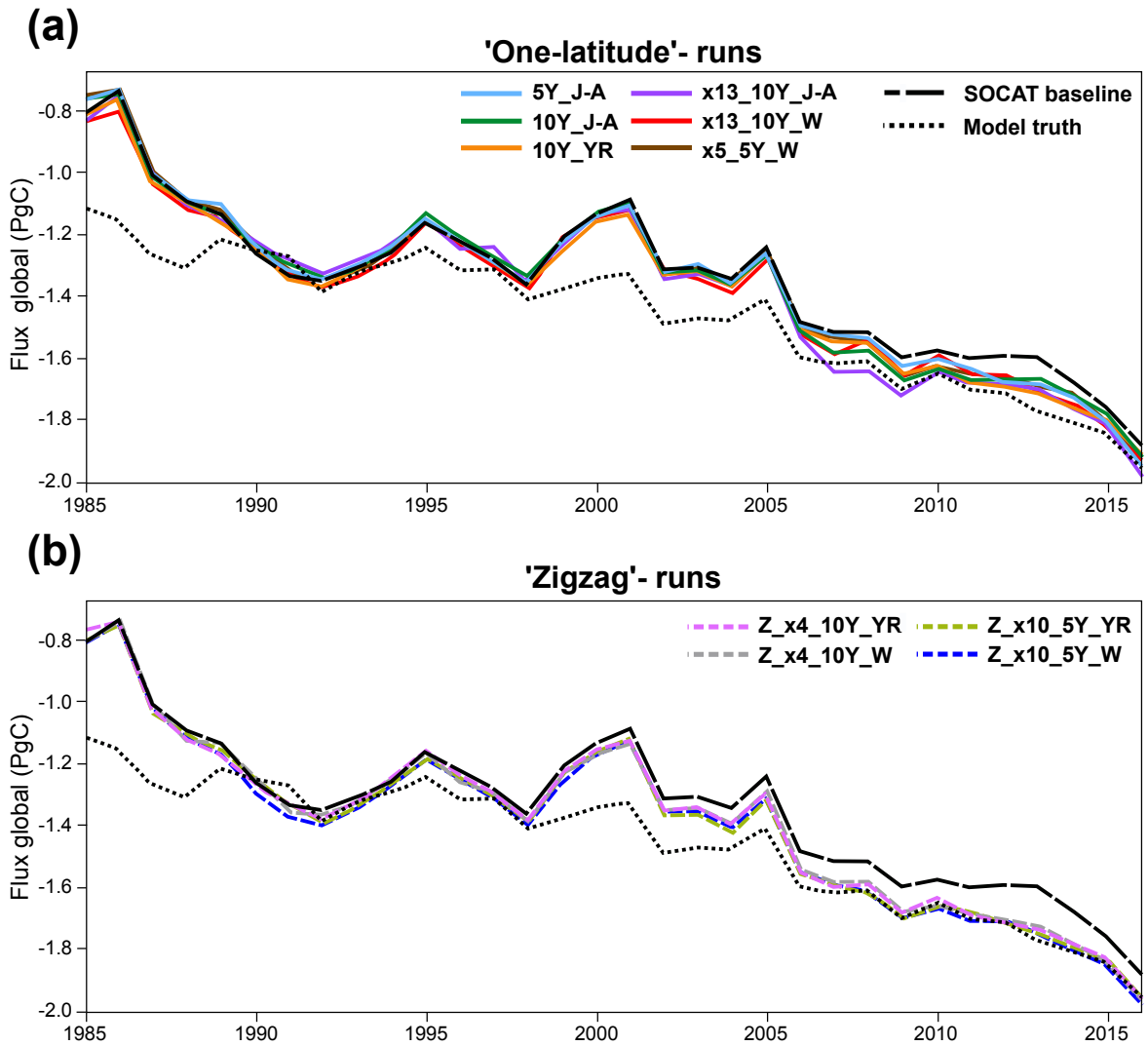

Figure S12. Globally annually averaged air-sea CO_2 flux for 'one-latitude' (**a**) and 'zigzag' (**b**) runs (see **Table 1**), including the SOCAT baseline (black dashed line) and the testbed 'model truth' (black dotted line), averaged over the 75 ensemble members. Runs highlighted in bold correspond to the two selected runs whose maps are shown in the main text (**Figs. 4, 6 7** and **9**). Generally, the Saildrone USV additions lead to an increased sink in the Southern Ocean compared to the SOCAT baseline. The 'zigzag' runs (dashed lines) generate a stronger sink compared to the 'one-latitude' runs (solid lines), and closely match the model truth the last ten years of the testbed.

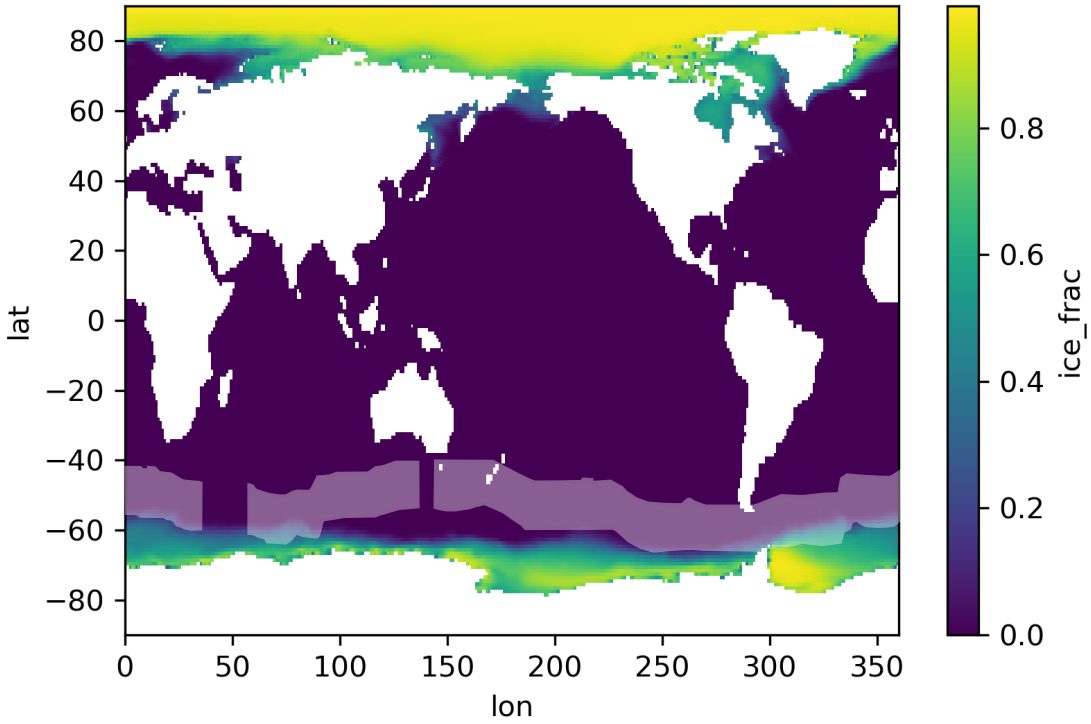

Figure S13: Map of global sea-ice extent as defined by the SeaFlux product (Fay et al., 2021). Light shaded areas represent the spatial extent of Saildrone USV sampling for run 'x13_10Y_J-A' and 'x13_10Y_W', where the 'real-world' Saildrone USV track (Sutton et al., 2021) was repeated six times by 1° north and south. The southernmost sampling occurs within the ice-zone as defined by ice fraction ('ice-frac') > 0.

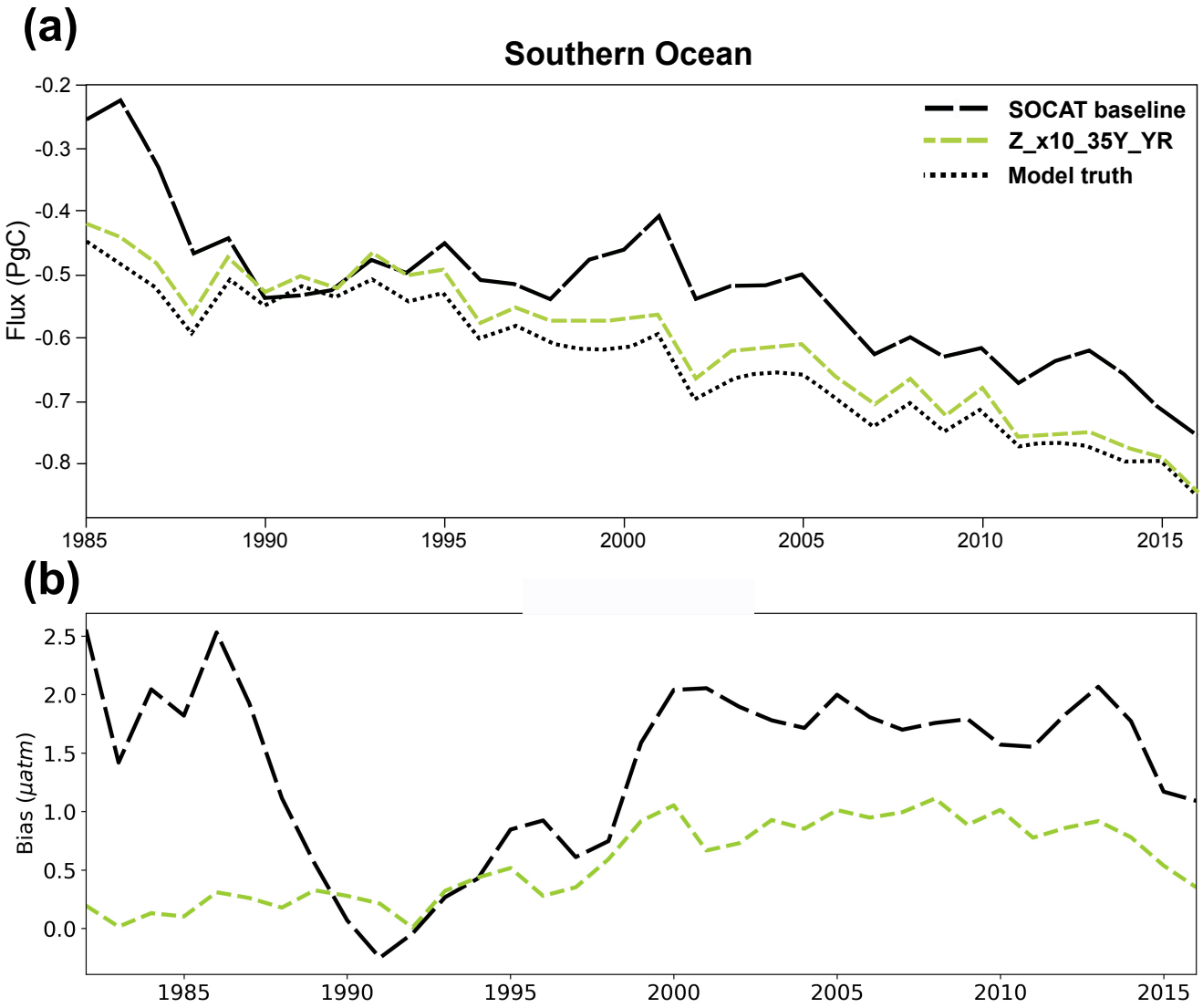

Figure S14. Southern Ocean (> 35° S) annually averaged air-sea CO_2 flux (**a**) and bias (**b**) for run 'Z_x10_35Y_YR' and the SOCAT baseline. Run 'Z_x10_35Y_YR' closely match the 'model truth' (black dotted line) air-sea flux for the entire testbed period (**a**), and show a significant dampening in bias variability compared to the SOCAT baseline (**b**).

Run name	Baseline	5Y_J-A	10Y_J-A	10Y_YR	x13_10Y_J-A	x13_10Y_W	x5_5Y_W	Z_x4_10Y_YR	Z_x4_10Y_W	Z_x10_5Y_YR	Z_x10_5Y_W
Saildrone track	NA	One-lat	One-lat	One-lat	One-lat	One-lat	One-lat	Zigzag	Zigzag	Zigzag	Zigzag
Years of sampling	NA	5	10	10	10	10	5	10	10	5	5
Duration of sampling	NA	Jan-Aug	Jan-Aug	Year-round	Jan-Aug	SO winter	SO winter	Year-round	SO winter	Year-round	SO winter
Additional observations	NA	**2,075**	**4,150**	**4,150**	**44,250**	**25,395**	**5,022**	**7,600**	**2,500**	**11,400**	**3,800**
Mean bias (µatm)											
Testbed period (1982-2016)											
Globally	0.63	0.59	0.59	0.52	0.53	**0.39**	0.57	0.51	0.51	0.45	0.44
NORTH (35°N-90°N)	**0.11**	0.24	0.20	0.25	0.20	0.17	0.16	0.16	0.16	0.12	0.20
MID (35°S-35°N)	0.23	0.21	0.22	0.14	0.20	0.15	0.23	0.20	0.18	**0.13**	0.18
SOUTH (90°S-35°S)	1.4	1.3	1.2	1.1	1.1	**0.80**	1.2	1.1	1.1	1.0	0.87
SO winter months (JJA)	1.3	1.2	1.2	1.1	1.1	**0.90**	1.2	0.93	1.0	0.94	0.95
SO summer months (DJF)	0.070	0.11	0.15	0.10	0.15	**0.019**	0.11	0.25	0.073	0.16	0.066
2006/2012-2016											
Globally	0.51*	0.27	0.34	0.28	0.19	**0.03**	0.21	0.23	0.24	0.17	0.07
SOUTH (90°S-35°S)	1.6*	0.93	1.1	1.0	0.72	**0.37**	0.73	0.89	0.92	0.67	0.55
SOUTH (90°S-35°S) Jun, Jul, Aug	4.2*	2.6	2.7	2.8	2.2	1.8	2.5	1.8	2.4	**1.2**	2.0
1984-1994	0.61	0.66	0.64	0.57	0.65	**0.50**	0.67	0.60	0.59	0.54	0.53
1995-2005	0.52	0.51	0.51	0.46	0.46	**0.35**	0.47	0.43	0.43	0.39	0.37
Mean RMSE (µatm)											
Testbed period (1982-2016)											
Globally	7.9	7.9	7.9	7.9	7.8	7.8	7.9	**7.7**	7.8	**7.7**	7.8
NORTH (35°N-90°N)	**9.1**	**9.1**	**9.1**	9.4	9.4	9.4	**9.2**	9.3	9.4	9.4	9.4
MID (35°S-35°N)	**7.2**	**7.2**	**7.2**	7.3	7.4	7.4	**7.2**	7.3	7.3	7.4	7.3
SOUTH (90°S-35°S)	8.5	8.4	8.4	8.5	8.4	8.3	8.3	8.2	8.4	**8.1**	8.3
2006/2012-2016											
Globally	7.7*	7.7	7.5	7.5	7.4	7.3	7.6	**7.2**	7.4	**7.2**	7.5
SOUTH (90°S-35°S)	8.4*	8.2	8.1	7.8	7.6	7.5	8.0	7.2	7.7	**6.9**	7.6
SOUTH (90°S-35°S) Jun, Jul, Aug	9.2*	8.4	8.4	8.2	7.6	7.5	8.3	7.5	7.8	**6.9**	7.5
1984-1994	8.3	8.3	8.4	8.3	8.3	8.3	8.3	**8.2**	8.3	**8.2**	**8.2**
1995-2005	7.6	7.6	7.6	7.6	7.6	7.5	7.6	**7.4**	7.5	7.5	7.5

Table S1: Mean bias and RMSE (in µatm) for various time periods, latitude bands and seasons for all runs, including the SOCAT baseline. Green, bold values represent the best score for each category. The 'one-latitude' ('one-lat') track incorporates the 'real-world' Saildrone USV route from Sutton et al. (2021), while the 'zigzag' track represents potential future meridional sampling (see **Fig. 2**; **Table 1**). 'Additional observations' = 1°x1° monthly Saildrone USV observations (not including SOCAT). J-A= January-August. YR = year-round. W = southern hemisphere winter. x4, x5, x10 and x13 = four, five, ten and 13 USVs. SO winter = Southern hemisphere winter months, i.e., June, July, August, and also including September. *Average value of the mean of 2006-2016 and 2012-2016.

Mean flux (Pg C/yr)	Model truth	SOCAT baseline
Globally		
1982-2016	-1.5	-1.3
2006-2016	-1.7	-1.6
2012-2016	-1.8	-1.7
SOUTH (90°S-35°S)		
1982-2016	-0.6	-0.5
2006-2016	-0.8	-0.7
2012-2016	-0.8	-0.7

Table S2: Global and Southern Ocean mean air-sea CO_2 fluxes (in Pg C yr^{-1}) for the testbed 'model truth' and the SOCAT baseline.

Run name	5Y J-A		10Y J-A		10Y YR		x13 10Y J-A		x13 10Y W		x5 5Y W		Z x10 5Y W		Z x4 10Y W		Z x10 5Y YR		Z x10 5Y W	
Saildrone track	One-lat		One-lat		One-lat		One-lat		One-lat		One-lat		Zigzag		Zigzag		Zigzag		Zigzag	
Years of sampling	5		10		10		10		10		5		10		10		5		5	
Duration of sampling	Jan-Aug		Jan-Aug		Shifted		Jan-Aug		SO winter		SO winter		Year-round		SO winter		Year-round		S winter	
Additional observations	**2,075**		**4,150**		**4,150**		**44,250**		**25,395**		**5,022**		**7,600**		**2,500**		**11,400**		**3,800**	
	% change	**diff**	**% change**	**diff**	**% change**	**diff**	**% change**	**diff**	**% change**	**diff**	**% change**	**diff**	**% change**	**diff**	**% change**	**diff**	**% change**	**diff**	**% change**	**diff**
Globally																				
1982-2016	-1	0.01	-1	0.02	-2	0.03	-2	0.03	-2	0.03	-1	0.02	-3	0.04	-3	0.04	-4	0.05	-4	0.06
2006/2012-2016	-4	0.06	-3	0.06	-4	0.06	-6	0.09	-3	0.06	-3	0.06	-5	0.09	-5	0.08	-6	0.10	-7	0.12
90°S-35°S																				
1982-2016	-2	0.01	-3	0.02	-5	0.03	-6	0.03	-5	0.03	-3	0.02	-8	0.04	-7	0.04	-9	0.05	-11	0.06
2006/2012-2016	-10	0.07	-9	0.06	-10	0.07	-14	0.09	-9	0.06	-9	0.06	-13	0.09	-13	0.08	-15	0.10	-15	0.10

Table S3: Difference ('diff') between calculated mean air-sea flux of individual Saildrone USV runs from the SOCAT baseline (in Pg C yr^{-1}), and change shown by %. There is a negative change from the SOCAT baseline, i.e., all Saildrone USV runs predict a stronger sink. 'Additional observations' = 1°x1° monthly Saildrone observations (not including SOCAT). 'SO winter' = Southern hemisphere winter months, i.e., June, July, August, and also including September. Testbed period = 1982-2016.