

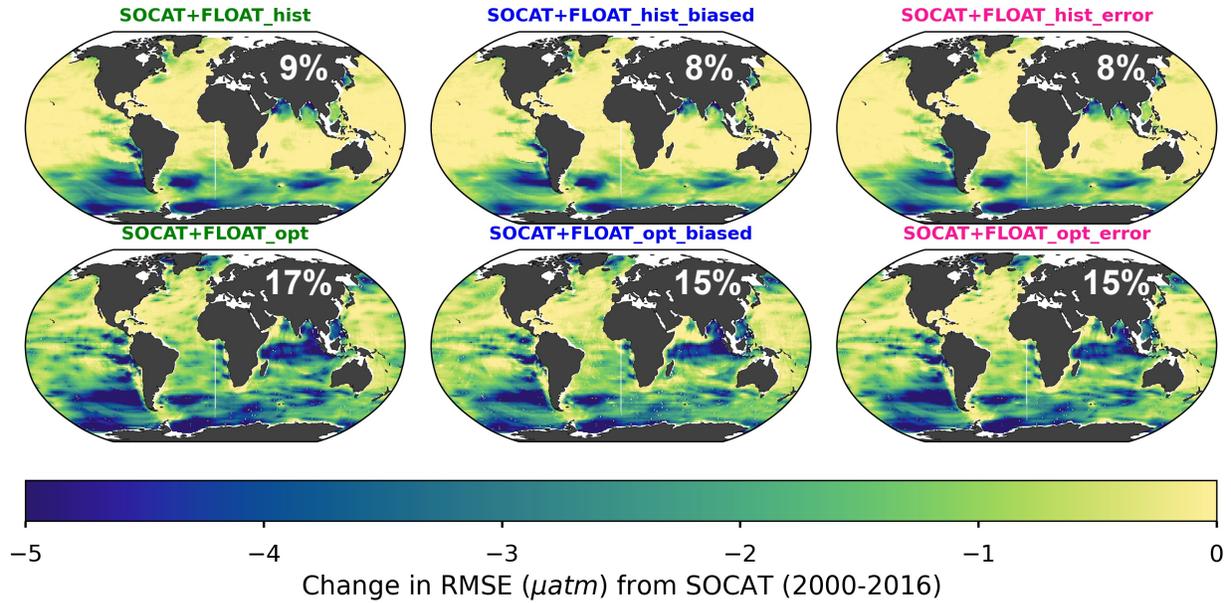
Supplementary Material

The importance of adding unbiased Argo observations to the ocean carbon observing system

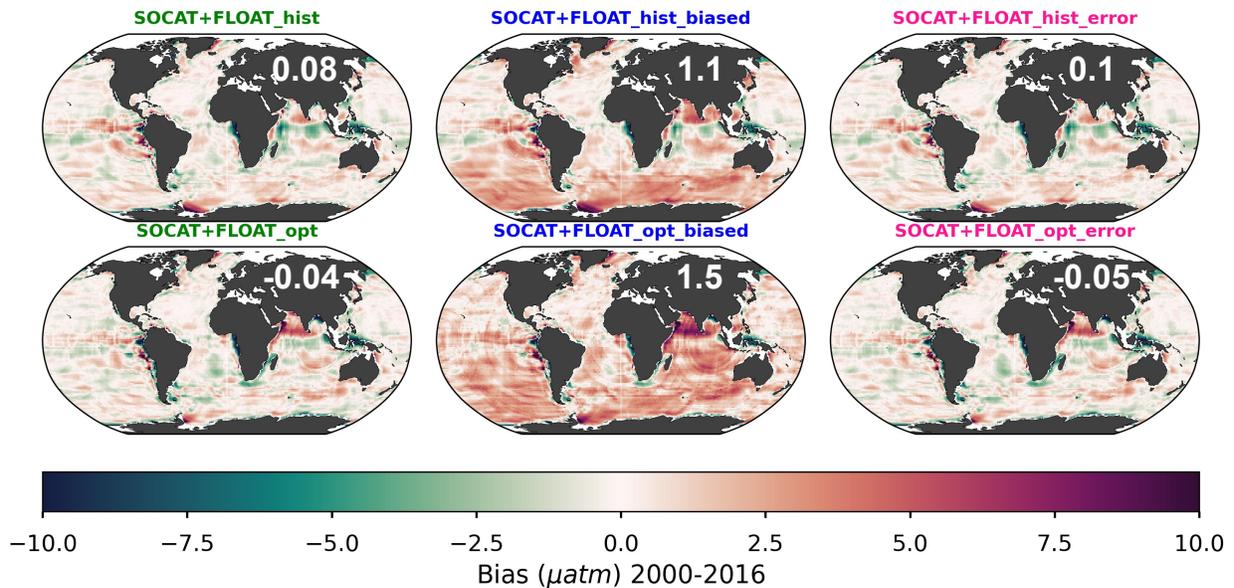
Thea H. Heimdal¹ & Galen McKinley¹

¹Columbia University and Lamont-Doherty Earth Observatory, Palisades, NY, USA

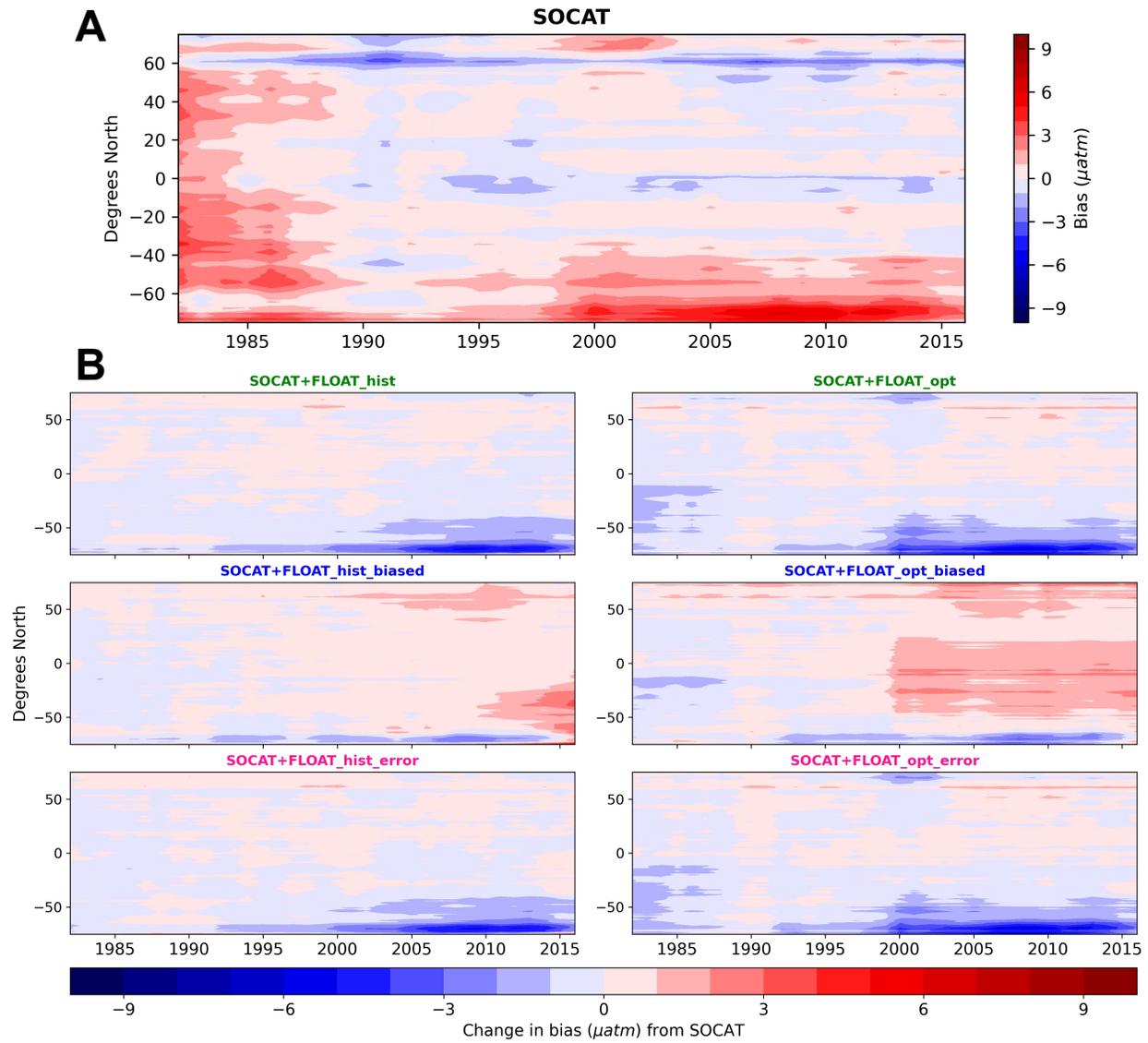
Corresponding author: theimdal@ldeo.columbia.edu



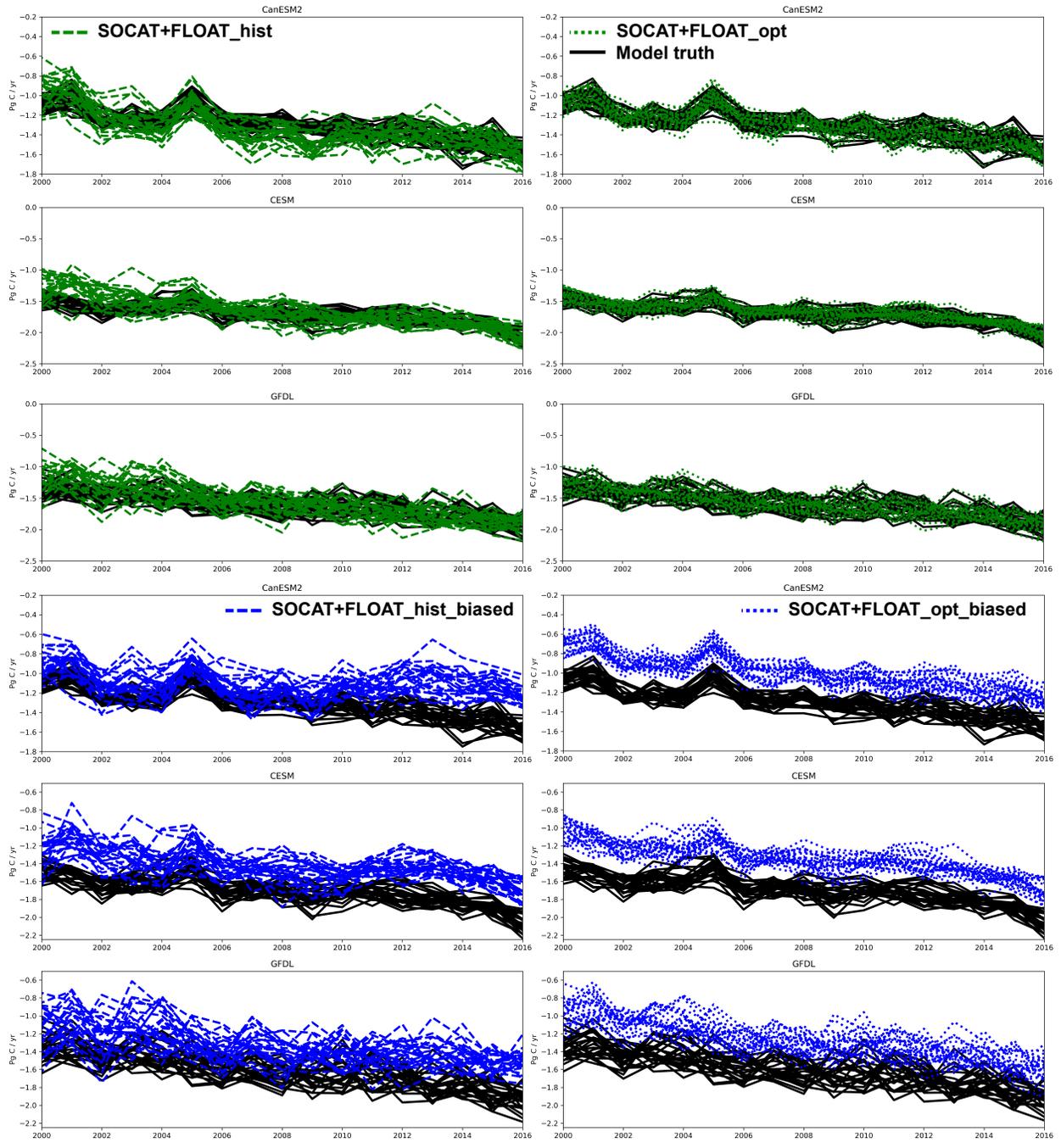
Supplementary Figure S1: Change in RMSE when comparing the six float experiments to the ‘SOCAT’ run. Values on panels represent improvement (in %) compared to the ‘SOCAT’ RMSE global mean for 2000-2016 (mean of all 75 members).



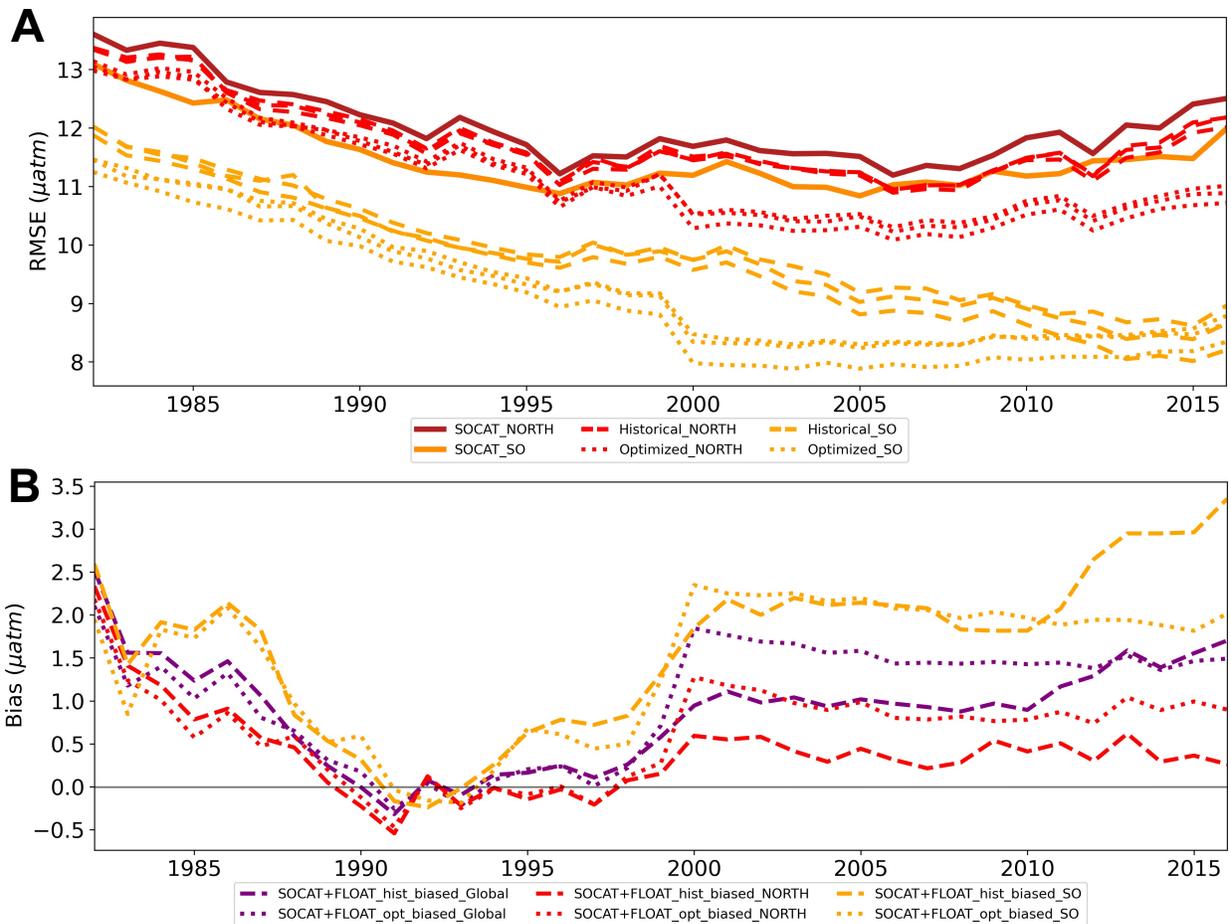
Supplementary Figure S2: Mean bias (all 75 members) for the six float experiments over the period of 2000 through 2016. Values on panels show the global mean (2000-2016) bias in μatm .



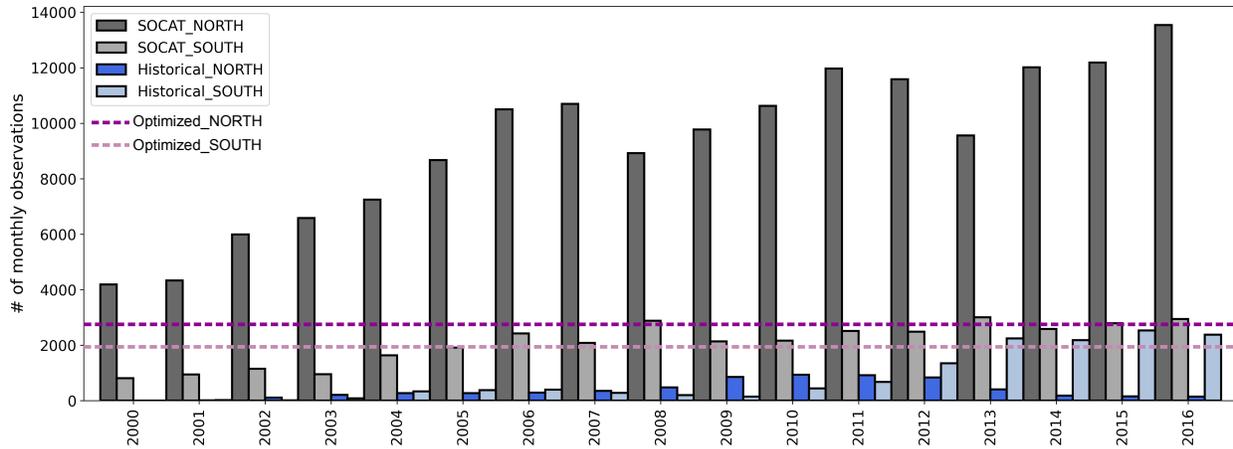
Supplementary Figure S3: Zonal mean, annual mean Hovmöller of bias (mean of all 75 members) for the ‘SOCAT’ run (A) and change in bias for the float experiments compared to the ‘SOCAT’ run (B).



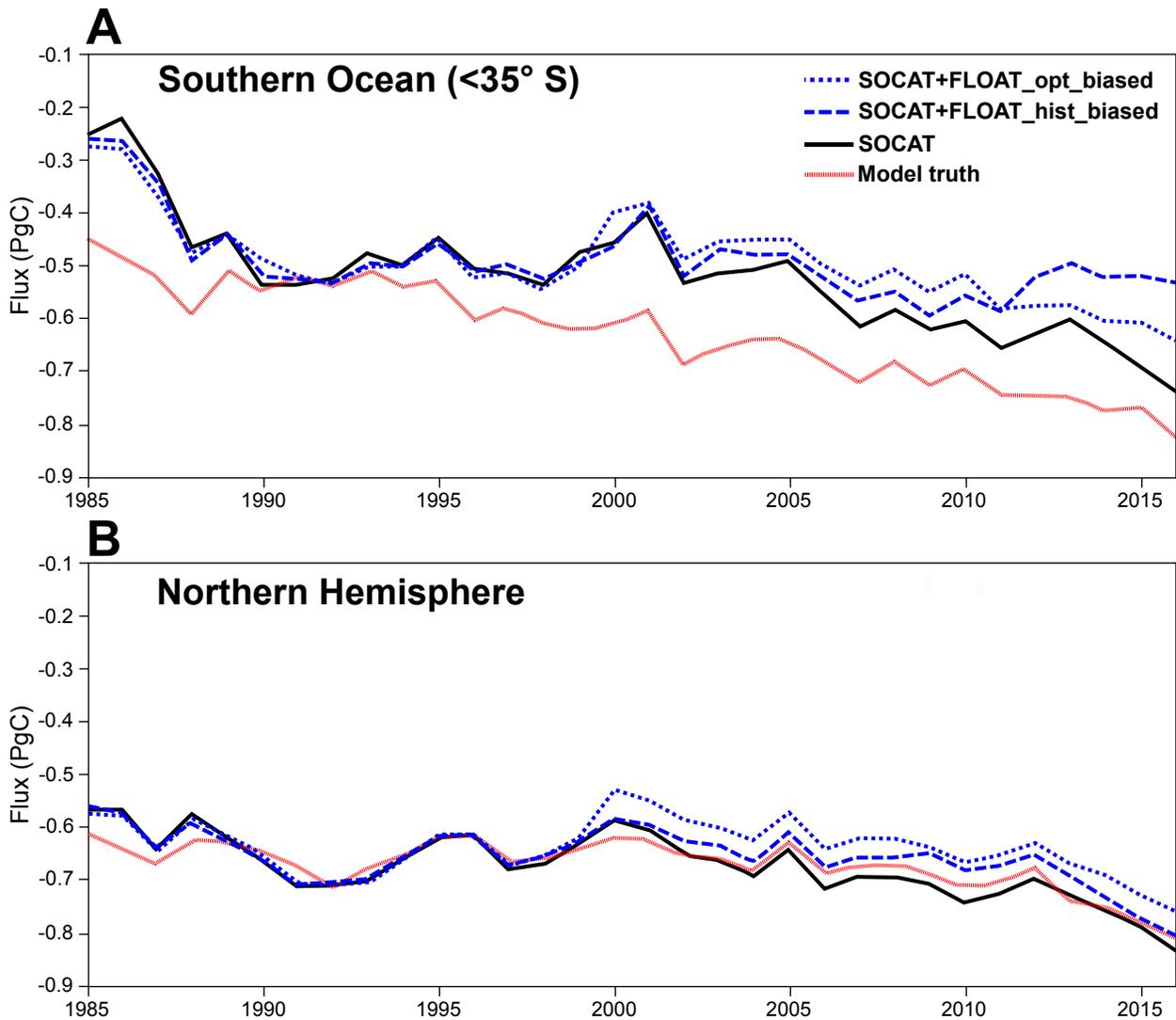
Supplementary Figure S4: Global annually averaged air-sea CO₂ flux for the ‘baseline’ (green) and ‘biased’ (blue) experiments compared to the ‘model truth’ (black) for the 25 members of the three models in the Large Ensemble Testbed.



Supplementary Figure S5: Annual mean RMSE (all 75 members) for all six float experiments and the ‘SOCAT’ run over the Southern Ocean ($< 35^\circ \text{S}$) and the northern hemisphere ($0\text{-}90^\circ \text{N}$) (A). Annual mean bias (all 75 members) for the ‘biased’ experiments, globally (purple), over the Southern Ocean ($< 35^\circ \text{S}$; yellow) and the northern hemisphere ($0\text{-}90^\circ \text{N}$; red) (B).



Supplementary Figure S6: Bar plot showing the number of $1^{\circ} \times 1^{\circ}$ monthly observations per year sampled from the Large Ensemble Testbed, for the ‘Historical’ floats (blue bars), ‘Optimized’ floats (purple/pink dashed lines) and the SOCAT database (gray bars) from the Southern Ocean (90°S - 35°S ; ‘SOUTH’; light colors) and the northern hemisphere (0 - 90°N ; ‘NORTH’; dark colors).



Supplementary Figure S7: Annually averaged (all 75 members) air-sea CO₂ flux over the Southern Ocean (< 35° S; **A**) and the northern hemisphere (0-90° N; **B**) for the biased float experiments, the 'SOCAT' run and the 'model truth'.

	SOCAT	Historical			Optimized		
		Baseline	Biased	Error	Baseline	Biased	Error
BIAS							
All 75 members							
<i>Global 2000-2016 mean</i>	0.6	0.08	1.1	0.1	-0.04	1.5	-0.05
1 IQR	0.5	0.4	0.3	0.3	0.1	0.2	0.2
Q1	0.9	0.3	1.3	0.3	0.05	1.6	0.03
Q3	0.4	-0.1	1.0	-0.1	-0.1	1.5	-0.1
CanESM2							
<i>Global 2000-2016 mean</i>	0.3	-0.1	0.9	-0.1	-0.1	1.4	-0.1
1 IQR	0.4	0.3	0.4	0.3	0.1	0.2	0.1
Q1	0.5	-0.01	1.1	0.02	-0.1	1.5	-0.05
Q3	0.1	-0.3	0.8	-0.3	-0.2	1.3	-0.1
CESM							
<i>Global 2000-2016 mean</i>	0.9	0.2	1.2	0.2	-0.1	1.6	-0.1
1 IQR	0.4	0.3	0.3	0.3	0.1	0.1	0.1
Q1	1.1	0.3	1.4	0.3	0.01	1.6	0.006
Q3	0.7	0.002	1.1	0.01	-0.09	1.5	-0.1
GFDL							
<i>Global 2000-2016 mean</i>	0.6	0.2	1.3	0.2	0.03	1.6	0.03
1 IQR	0.4	0.2	0.1	0.2	0.1	0.1	0.2
Q1	0.9	0.3	1.3	0.3	0.1	1.6	0.1
Q3	0.5	0.1	1.2	0.2	-0.01	1.5	-0.1
RMSE							
All 75 members							
<i>Global 2000-2016 mean</i>	11.6	10.5	10.7	10.6	9.6	9.8	9.8
1 IQR	2.1	2.3	2.2	2.3	2.5	2.4	2.5
CanESM2							
<i>Global 2000-2016 mean</i>	12.9	12.0	12.2	12.1	11.3	11.4	11.5
1 IQR	0.3	0.3	0.2	0.2	0.3	0.2	0.3
CESM							
<i>Global 2000-2016 mean</i>	11.1	9.7	9.9	9.8	8.6	8.9	9.0
1 IQR	0.4	0.2	0.2	0.2	0.2	0.1	0.2
GFDL							
<i>Global 2000-2016 mean</i>	10.6	9.6	9.8	9.8	8.6	8.9	8.9
1 IQR	0.3	0.1	0.2	0.2	0.1	0.2	0.2

Table S1. Overview of global mean (2000-2016) bias and RMSE and the interquartile range (IQR) (in μatm) for the full 75-member Large Ensemble Testbed and individual models. Bold values represent the “best” RMSE or bias score for each category; i.e., the lowest RMSE value, or bias value closest to zero.

2000-2016 annually averaged mean flux (Pg C/year)														
	SOCAT	Diff.	Historical						Optimized					
			Baseline	Diff.	Biased	Diff.	Error	Diff.	Baseline	Diff.	Biased	Diff.	Error	Diff.
Globally														
All 75 members	-1.4	-0.1	-1.5	-0.02	-1.3	-0.26	-1.5	-0.02	-1.5	-0.01	-1.2	-0.32	-1.5	-0.01
CESM	-1.5	-0.2	-1.7	-0.04	-1.4	-0.3	-1.7	-0.1	-1.7	-0.02	-1.4	-0.3	-1.7	-0.02
GFDL	-1.5	-0.1	-1.6	-0.04	-1.4	-0.3	-1.6	-0.04	-1.6	-0.01	-1.3	-0.3	-1.6	-0.01
CanESM2	-1.3	-0.04	-1.3	0.03	-1.1	-0.2	-1.3	0.03	-1.3	0.003	-1.0	-0.3	-1.3	0.003
All 75 members														
Northern Hemisphere	-0.7	0.01	-0.7	0.004	-0.7	-0.03	-0.7	0.004	-0.7	-0.002	-0.6	-0.1	-0.7	-0.001
Southern Ocean	-0.7	-0.1	-0.7	-0.02	-0.5	-0.2	-0.7	-0.02	-0.7	-0.01	-0.5	-0.2	-0.7	-0.01

Table S2. Overview of annually averaged mean (2000-2016) air-sea CO₂ flux (all 75 members) for the full 75-member Large Ensemble Testbed and individual models, globally, over the Southern Ocean (> 35° S) and the northern hemisphere (0-90° N). ‘Diff.’ = difference between the mean 2000-2016 ‘model truth’ and reconstructed air-sea flux for each experiment. Bold values represent the “best score” for each category (i.e., closest to ‘model truth’).