Supplementary Information – Table of Contents

Figures

Tables

Figure 1 Anomaly correlation for 2-meter temperature for retrospective forecast (1981-2016) lead times June- February and ERA 5. Blues indicate a positive anomaly correlation coefficient and good skill in the GEOS-S2S at predicting temperature and reds indicate poor skill/misleading forecasts. The stippling indicates regions of statistical significance (significance level of 0.05).

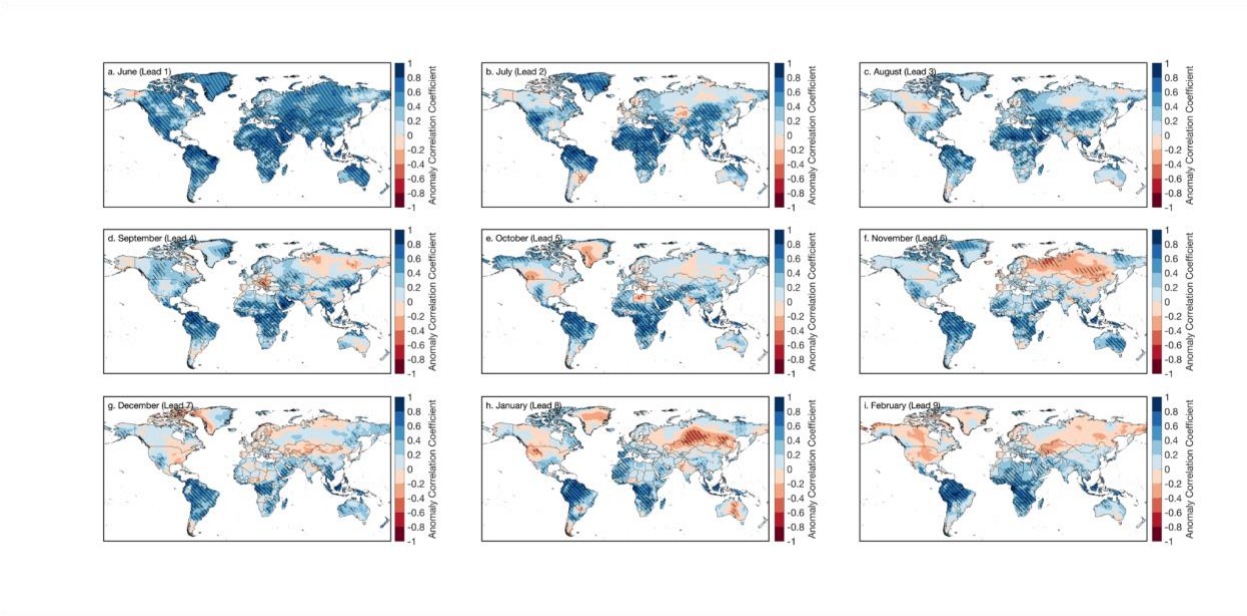


Figure 2. Anomaly root mean squared error between the 2-meter temperature for retrospective forecasts and ERA 5 (1981-2016) for lead months 1-3 (June, July, August). (a-c) All years. (d-f) Years with a sea surface temperature anomaly of less than 0.5°C in the Niño 3.4 region during forecast initialization (years exclusive of 1982, 1983, 1987, 1992, 1993, 1997, 1998, 2015). (g-i) years with a sea surface temperature anomaly of 0.5°C or higher in the Niño 3.4 region during forecast initialization (1982, 1983, 1987, 1992, 1993, 1997, 1998, 2015).

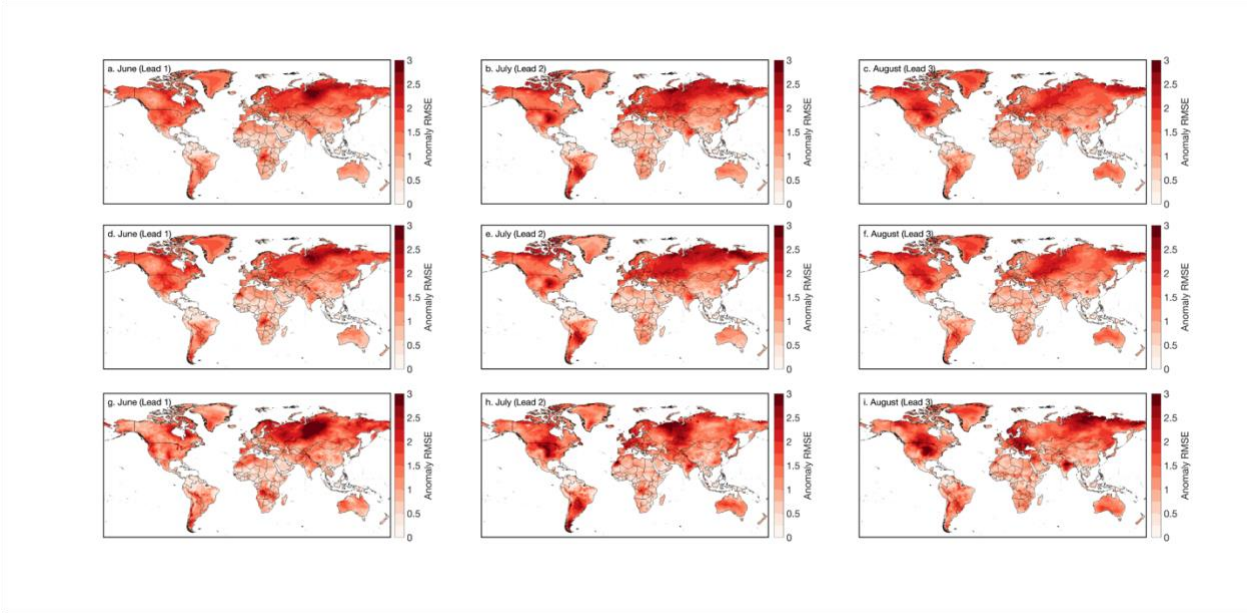


Figure 3. The mean number of ensemble members from all retrospective forecasts (1981-2016) that correctly characterize the 2-m air temperature anomaly into equiprobable categories (below normal, normal, and above normal) when assessed against ERA5 2-m air temperatures. Blues indicate that, on average, two or more of the four ensemble members accurately predict the ERA-5 2-m air temperature category and suggest the forecasts have higher confidence in the 2-m air temperature forecast. Reds and oranges indicate that, on average, less than two of the 4 ensemble members accurately predict the ERA-5 2-m air temperature category.

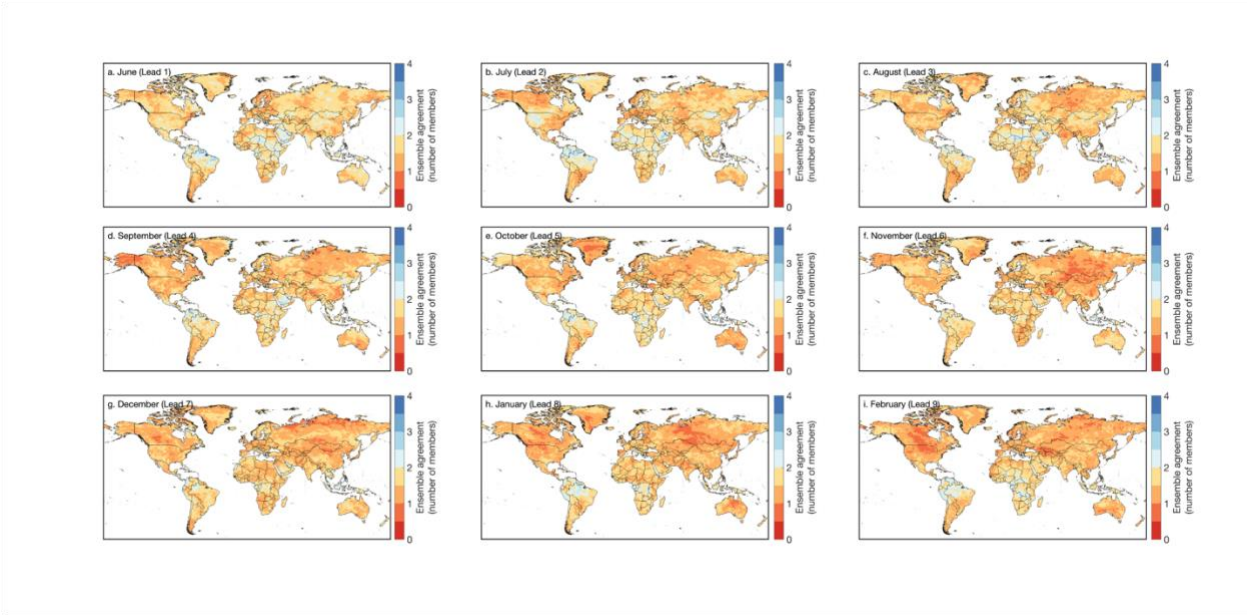


Figure 4. The mean number of ensemble members from retrospective forecasts initialized during neutral or cold ENSO conditions (years exclusive of 1982, 1983, 1987, 1992, 1993, 1997, 1998, 2015) that correctly characterize the 2-m air temperature anomaly into equiprobable categories (below normal, normal, and above normal) when assessed against ERA5 2-m air temperatures. Blues indicate that, on average, two or more of the four ensemble members accurately predict the ERA-5 2-m air temperature category and suggest the forecasts have higher confidence in the 2-m air temperature forecast. Reds and oranges indicate that, on average, less than two of the 4 ensemble members accurately predict the ERA-5 2-m air temperature category.

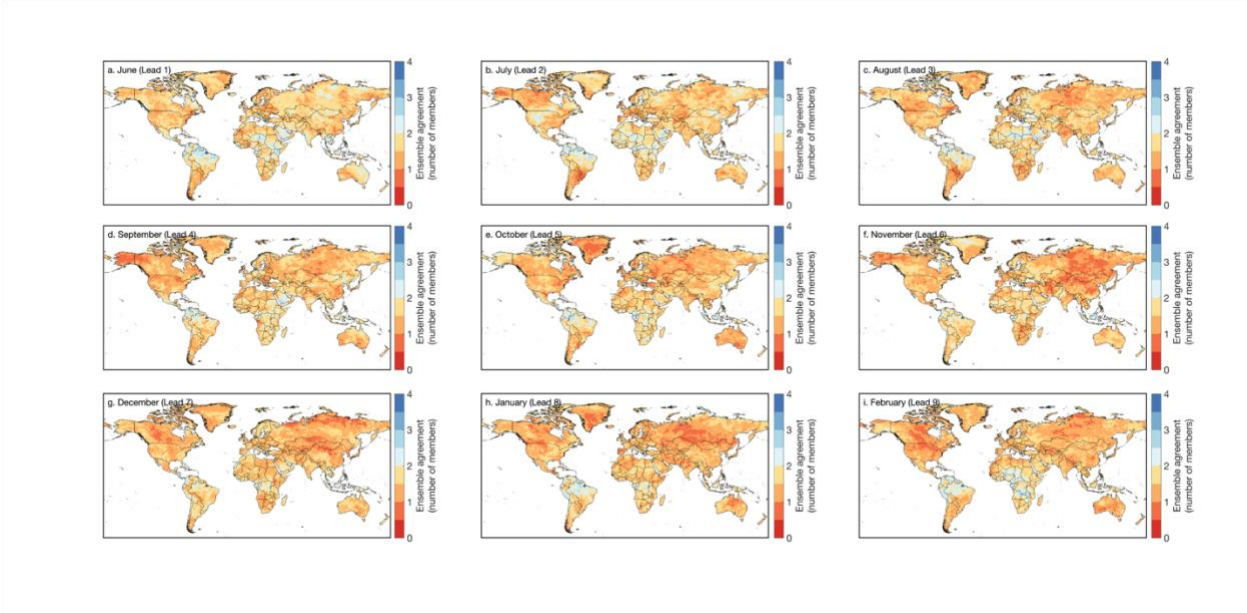


Figure 5. Anomaly correlation for 2-meter temperature for near real time forecast (2017-2024) lead months June to February and ERA 5. Blues indicate a positive anomaly correlation coefficient and good skill in the GEOS-S2S at predicting temperature and reds indicate poor skill/misleading forecasts. The stippling indicates regions of statistical significance (significance level of 0.10). Significance level reduced due to the limited number of near real-time forecast years.

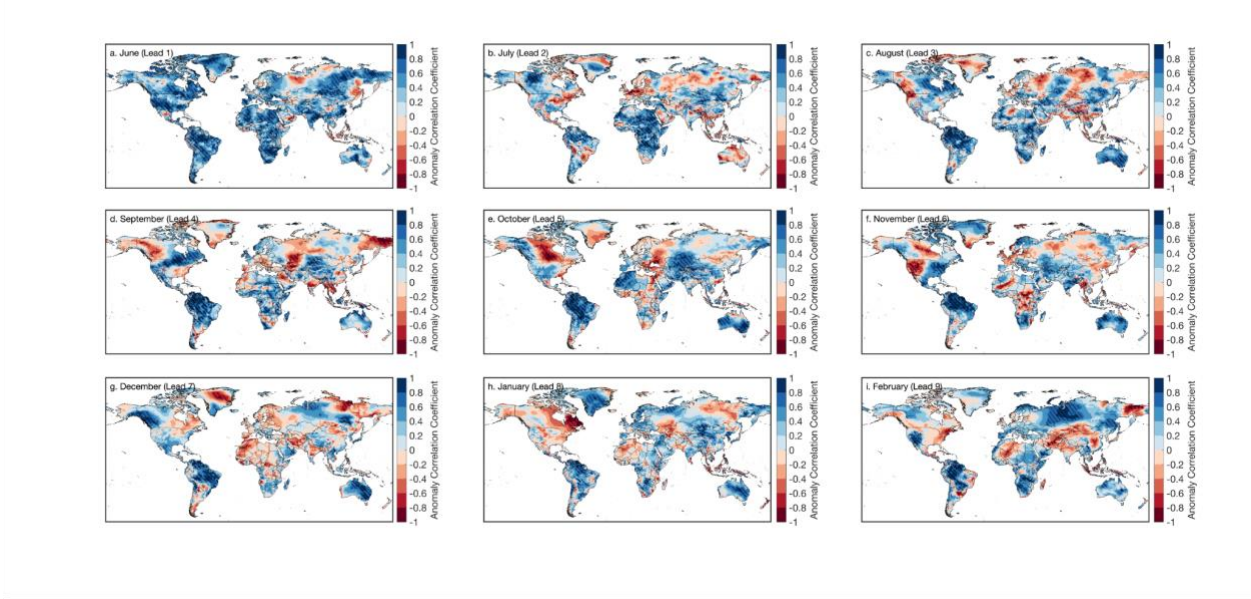


Figure 6. Ranked probability skill score (de-biased) for the near real time GEOS-S2S forecasts and ERA 5. Darker reds indicate a more beneficial forecast. Only skillful correlations are shown; light blues indicate regions where RPSS is zero or negative (misleading or no forecast skill).

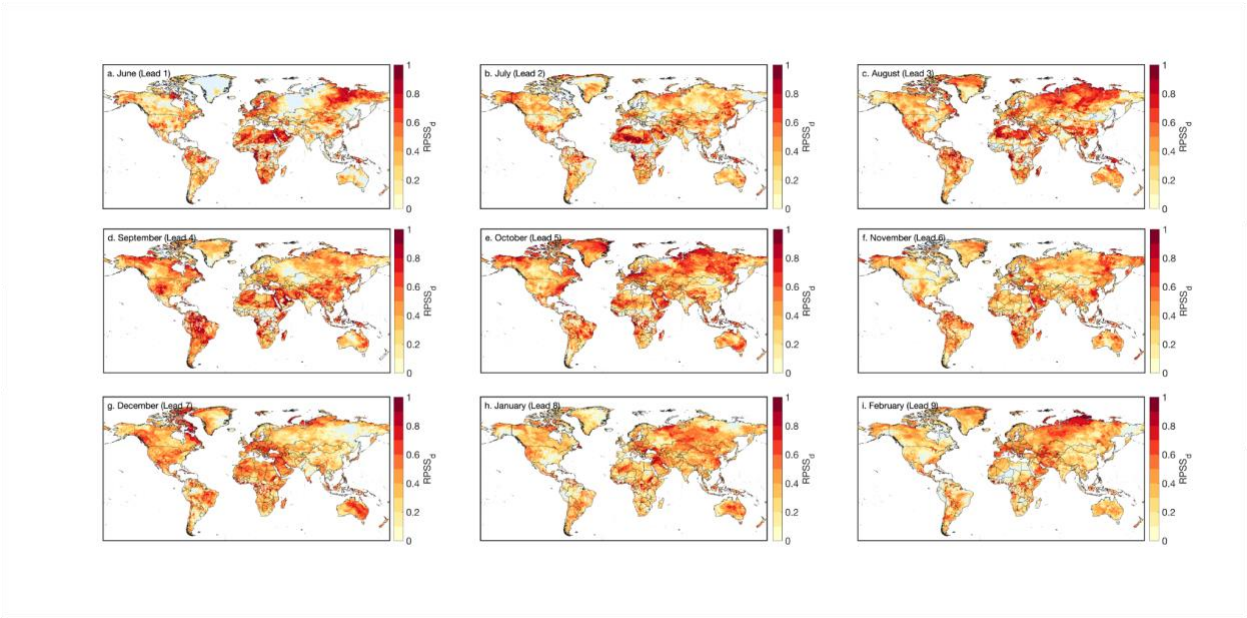


Figure 7. Spatially averaged anomaly RMSE of 2-m air temperatures for GEOS-S2S forecasts initialized in May 2024 and ERA 5. The spatial mean is taken over land for a range of latitude zones, including, 60N-60S (non-polar land), 60N-30N (Northern Hemisphere extra tropics), 30N-30S (tropics), and 30S-60S (Southern Hemisphere extra tropics).

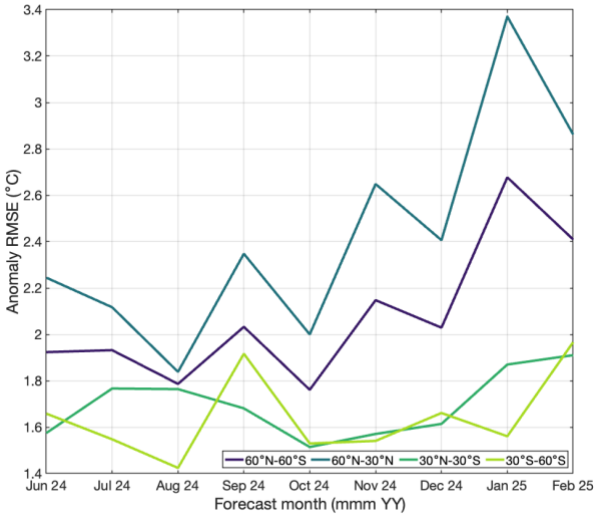


Figure 8. Anomaly correlation for 2-meter temperature for retrospective forecast (1981-2016) lead times June- February and ERA 5 for years with a sea surface temperature anomaly of less than 0.5°C in the Niño 3.4 region during forecast initialization (years exclusive of 1982, 1983, 1987, 1992, 1993, 1997, 1998, 2015). Blues indicate a positive anomaly correlation coefficient and good skill in the GEOS-S2S at predicting temperature and reds indicate poor skill/misleading forecasts. The stippling indicates regions of statistical significance (significance level of 0.05).

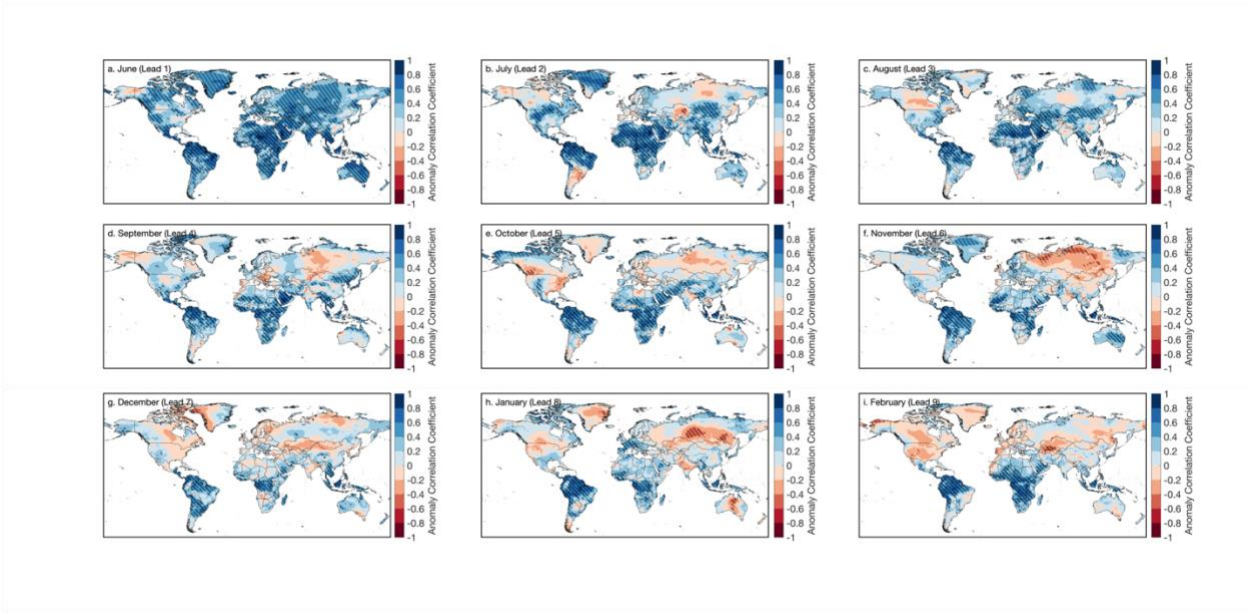


Figure 9 Anomaly correlation for 2-meter temperature for retrospective forecast (1981-2016) lead times June- February and ERA 5 for years with a sea surface temperature anomaly of 0.5°C or higher in the Niño 3.4 region during forecast initialization (1982, 1983, 1987, 1992, 1993, 1997, 1998, 2015). Blues indicate a positive anomaly correlation coefficient and good skill in the GEOS-S2S at predicting temperature and reds indicate poor skill/misleading forecasts. The stippling indicates regions of statistical significance (significance level of 0.05).

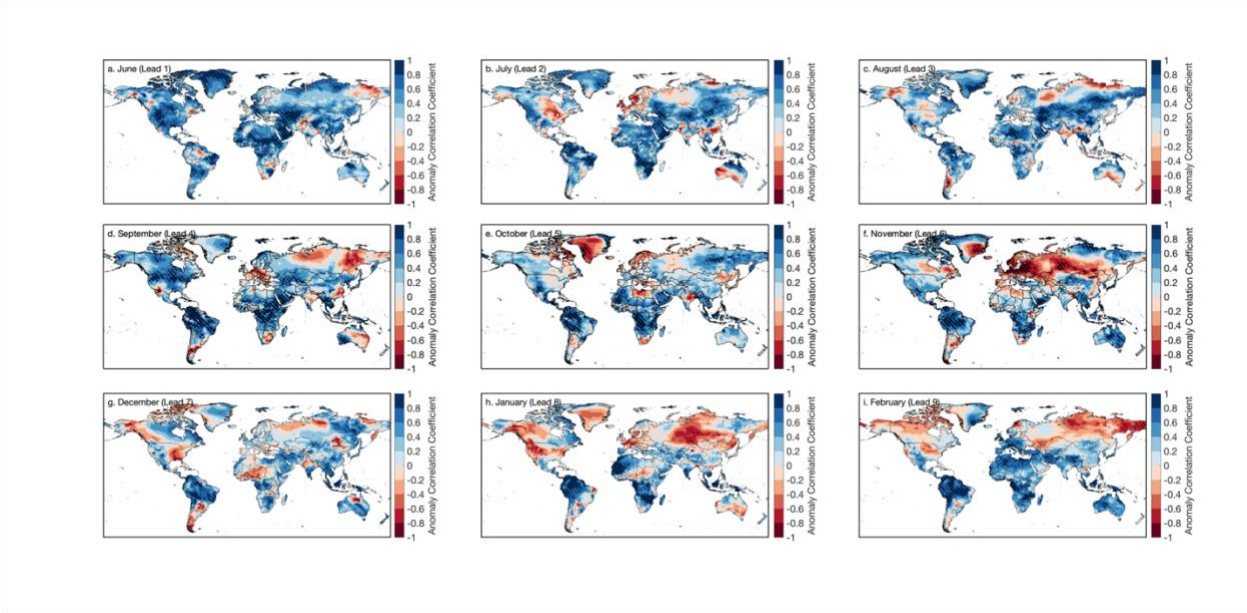


Figure 10. The mean number of ensemble members from retrospective forecasts initialized during warm ENSO conditions (1982, 1983, 1987, 1992, 1993, 1997, 1998, 2015) that correctly characterize the 2-m air temperature anomaly into equiprobable categories (below normal, normal, and above normal) when assessed against ERA5 2-m air temperatures. Blues indicate that, on average, two or more of the four ensemble members accurately predict the ERA-5 2-m air temperature category and suggest the forecasts have higher confidence in the 2-m air temperature forecast. Reds and oranges indicate that, on average, less than two of the 4 ensemble members accurately predict the ERA-5 2-m air temperature category.

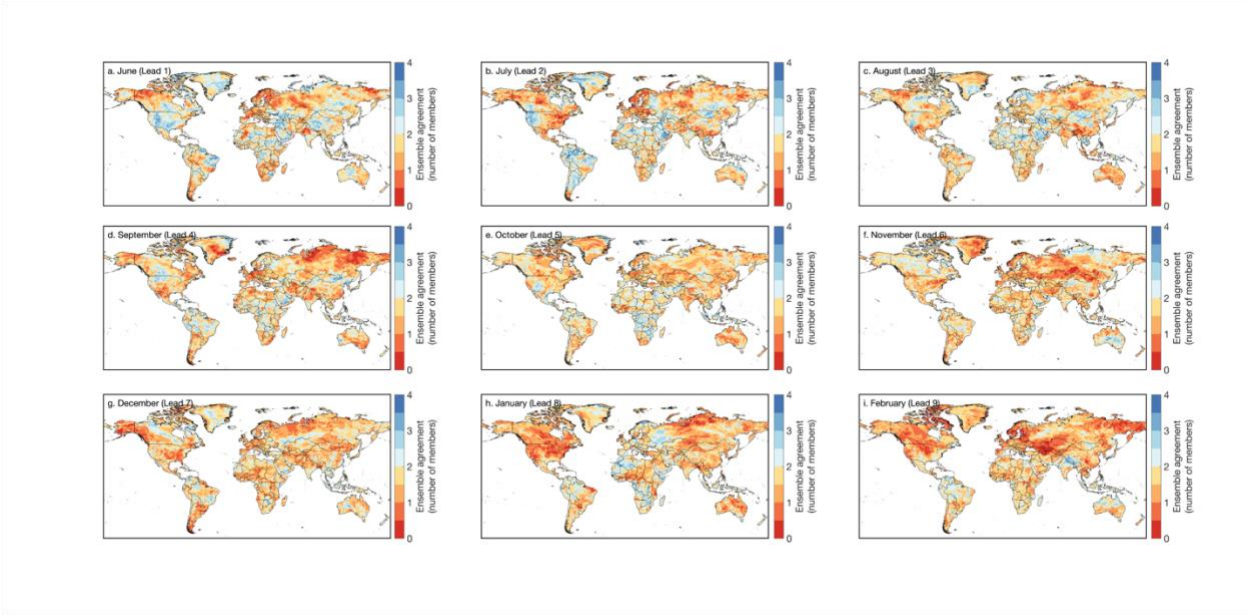
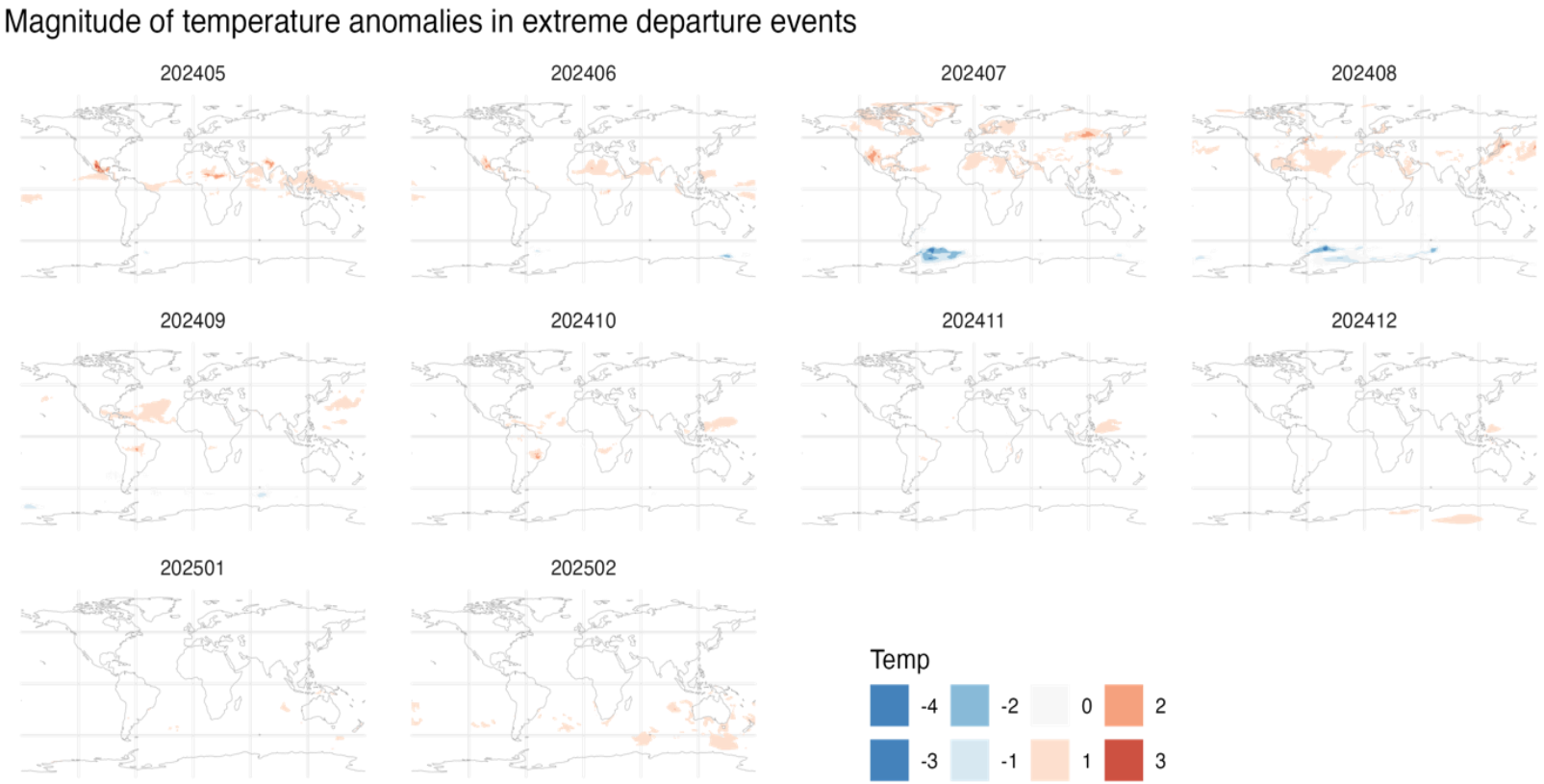


Figure 11. Season-2-season forecast anomalies in climate extreme departure events – those scoring historical (1980-2023) below or above record temperatures.



183
184

Figure 12. Comparison of species exposure thermal departure (average difference between species thermal limit and forecasted temperature) with subseasonal-to-seasonal climate (s2s) mean ensemble member and ERA5 reanalysis. Delta shows thermal differences median across forecasted dates.

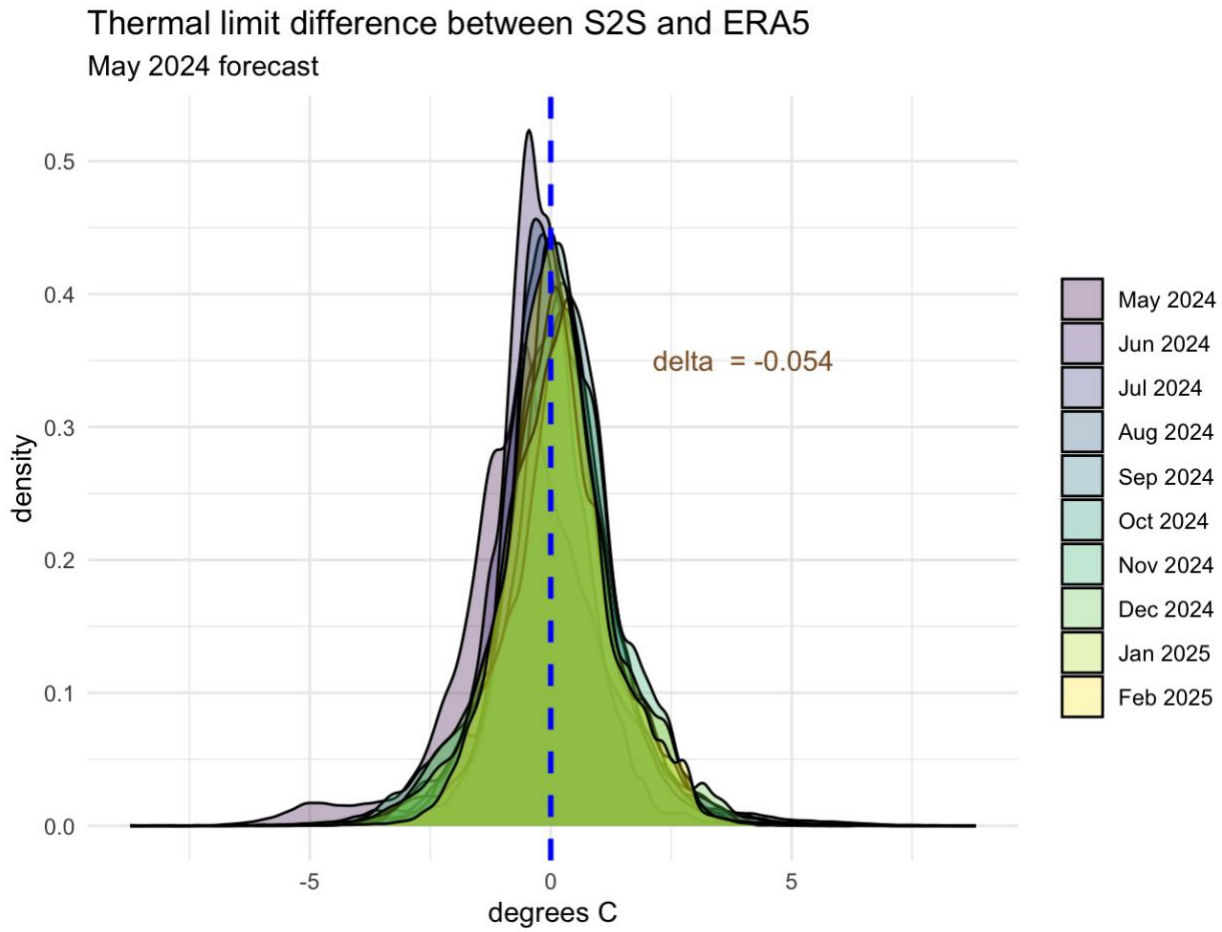
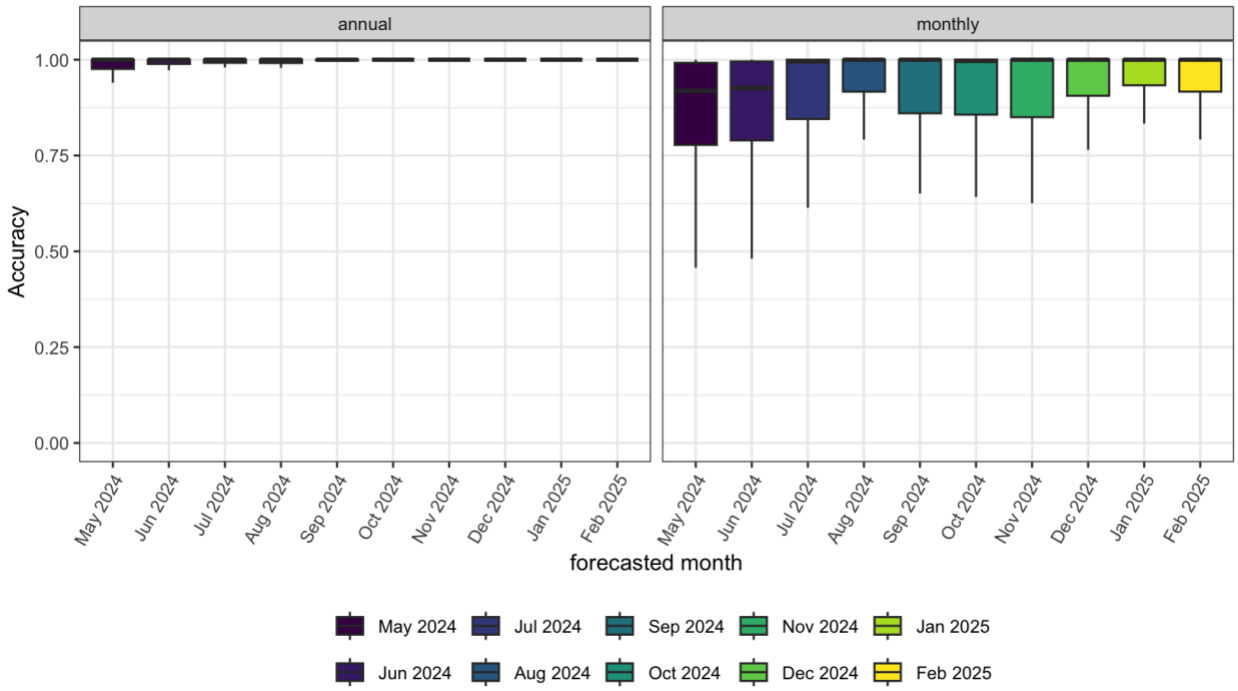


Figure 13. Species s2s exposure accuracy (correctly classified rate, as compared the ERA5 reanalysis product) for annual and monthly exposure and the first three model predictions. Outliers not shown for visual display



.

Figure 14. Sensitivity of GEOS-S2S prediction compared to the ERA5 reanalysis product for extreme and monthly exposure and the first three model predictions. Outliers not shown for visual display

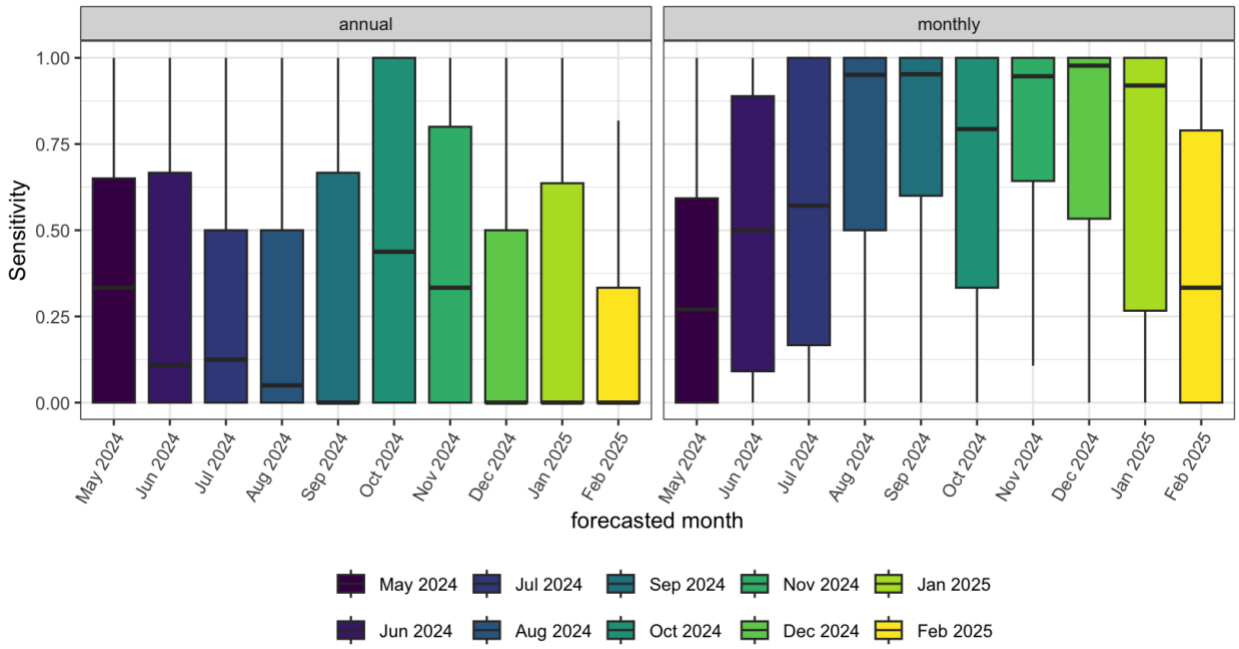


Figure 15. Probabilistic evaluation of exposure forecasts against S2S exposure using Brier Scores (A), and against the baseline (expected probabilistic outcome) through Brier Skill Scores (B). In B) dots denote medians. In A) outliers were visually excluded from plots (but not from data) for better interpretation of the graph

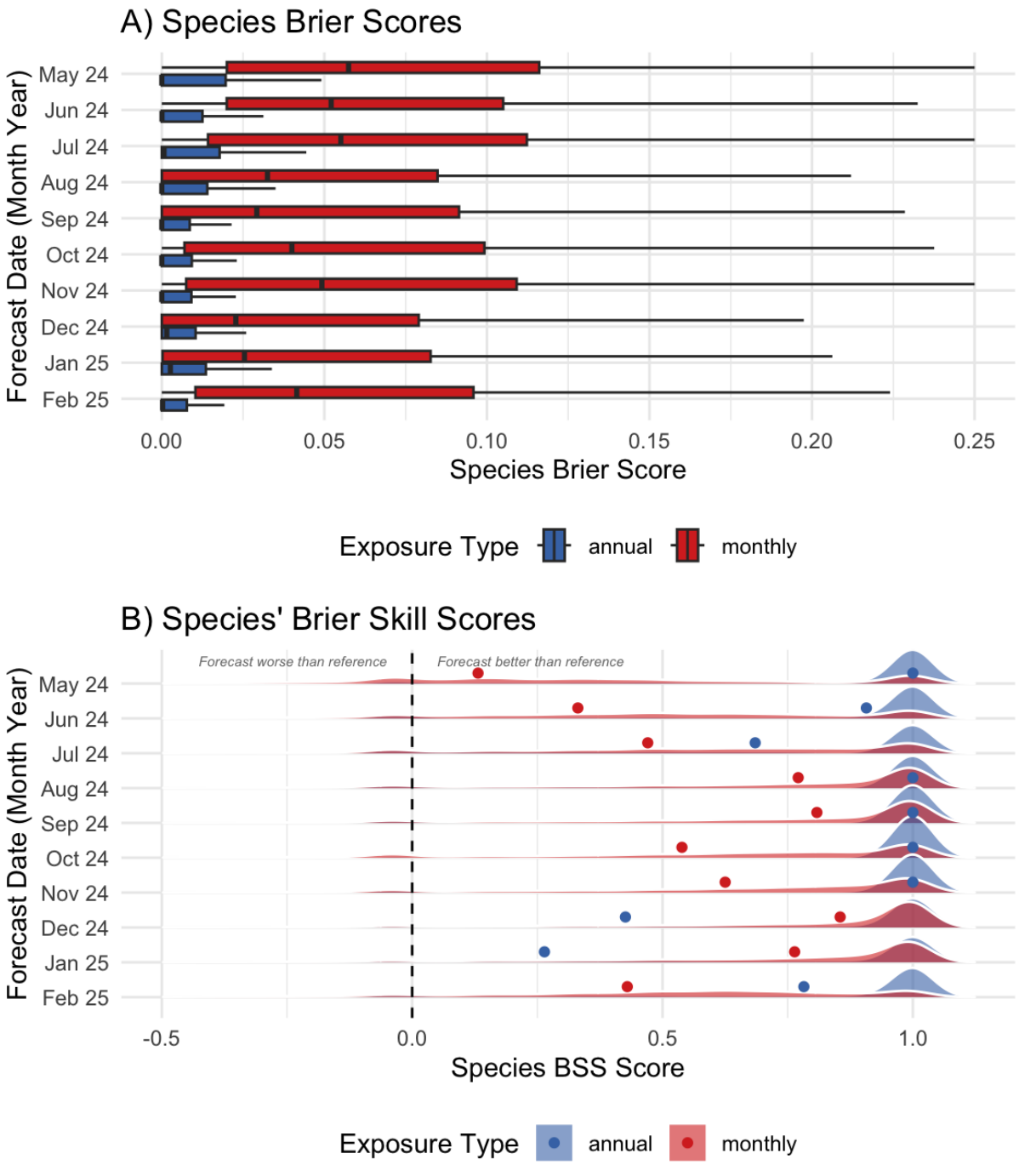
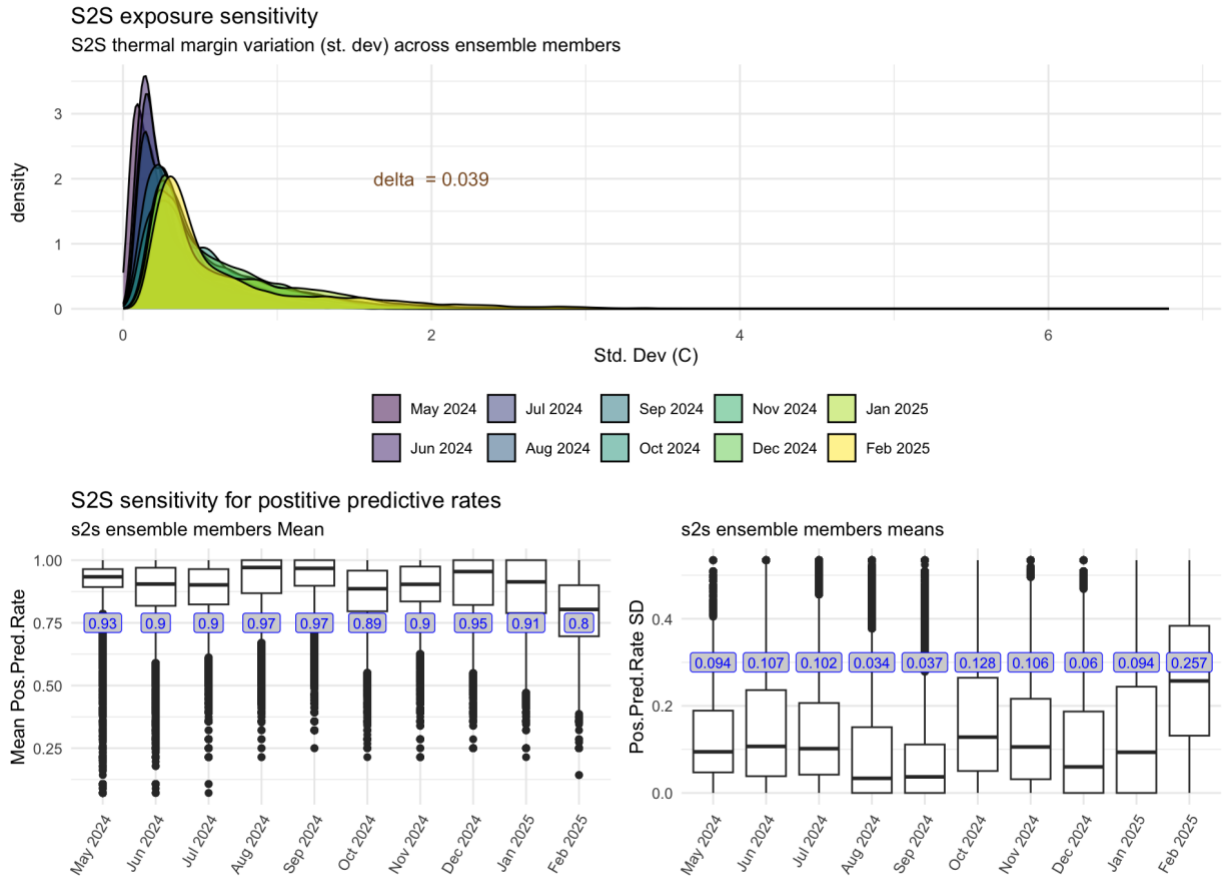


Figure 16. Exposure forecasting spread across s2s ensemble members predictions for the thermal margin and for the positive predicted rates. Leadtime means the Nth month in which the forward prediction is done. Leadtime 0 is May 2024, Leadtime 1 refers to June 2024, and Leadtime 2 refers to July 2024.



Supplementary Table 1 Distribution statistics for range size (km^2) by studied group. Q5, Q50 and Q95 show percentiles 5%, 50% and 100%.

Group	Q5	Mean	Q50	Q95
amphibians	2949.988	312146.3	23711.91	1610841
birds	14323.196	1979214.4	573553.27	8674543
mammals	6072.305	1477292.4	269974.22	7357987
reptiles	2978.450	607803.2	62212.03	3215404

Table 2. Weak relationship (Kendall correlation) between species range size and number of exposed sites, by type of exposure (annual and monthly, and forecast date in year-month format).

Forecast Date (yyyymm)	Kendall correlation	Exposure type
202405	-0.01	annual
202406	0.06	annual
202407	0.14	annual
202408	0.03	annual
202409	-0.03	annual
202410	-0.01	annual
202411	-0.02	annual
202412	0.00	annual
202501	-0.01	annual
202502	-0.02	annual
202405	0.01	monthly
202406	0.04	monthly
202407	0.12	monthly
202408	0.01	monthly
202409	0.00	monthly
202410	-0.02	monthly
202411	-0.07	monthly
202412	-0.08	monthly
202501	-0.09	monthly
202502	-0.06	monthly

Table 3. Variable entropy weights for ranking regional exposure according to different schemes. The *All-in* scenario considers all variables, the *Extreme first* scenario includes annual exposure variables only, the *Species-at-risk* variables subsets only to variables for vulnerable species and data deficient species (sensu IUCN), and the *Ecosystems-at-risk* approach takes into account exposure for widespread species (covering more than 50% of the region area). Direction means the way in which the variable prioritizes a given region: max means the higher the variable the region gets a higher prioritization, and the opposite for min. A value of 0 in a variable means that the variable is strongly correlated (r>0.7) with other variables and has not been considered for entropy weights, and a weight of NA implies that the variable has not been considered in a given approach.

Variable	Direction	All-in	Extreme first	Species-at-risk	Ecosystems-at-risk
Total area of Extreme exposure	max	1.5e-01	0.24	NA	NA
Total area of Monthly exposure	max	9.4e-02	NA	NA	NA
Proportional area of Extreme exposure	max	0.0e+00	0.00	NA	NA
Proportional area of Monthly exposure	max	0.0e+00	NA	NA	NA
Proportional area of Extreme exposure by widespread species	max	1.4e-01	0.21	NA	0.30
Proportional area of Monthly exposure by widespread species	max	4.5e-02	NA	NA	0.10
Number of widespread species exposed (extreme)	max	2.8e-01	0.43	NA	0.61
Number of widespread species exposed (month)	max	0.0e+00	NA	NA	0.00
Proportion of widespread species exposed (extreme)	max	0.0e+00	0.00	NA	0.00
Proportion of widespread species exposed (month)	max	0.0e+00	NA	NA	0.00
Proportional area of Monthly exposure by vulnerable species	max	0.0e+00	0.00	NA	NA
Proportional area of Monthly exposure by vulnerable species	max	0.0e+00	NA	NA	NA
Proportion of vulnerable species exposed (extreme)	max	0.0e+00	0.00	0.00	NA
Proportion of vulnerable species exposed (monthly)	max	0.0e+00	NA	0.00	NA
Proportion of vulnerable species exposed (extreme) to >50% of their range	max	0.0e+00	0.00	0.29	NA
Proportion of vulnerable species exposed (monthly) to >50% of their range	max	0.0e+00	NA	0.15	NA
Number of vulnerable species exposed (extreme) to >50% of their range	max	0.0e+00	0.00	0.36	NA
Number of vulnerable species exposed (monthly) to >50% of their range	max	0.0e+00	NA	0.00	NA
Number of vulnerable species exposed (extreme)	max	0.0e+00	0.00	0.00	NA
Number of vulnerable species exposed (monthly)	max	0.0e+00	NA	0.20	NA
Average extreme exposure incidence (N events)	max	0.0e+00	0.00	NA	NA
Average monthly exposure incidence (N events)	max	1.5e-01	NA	NA	NA
Total extreme exposure incidence (N events)	max	0.0e+00	0.00	NA	NA
Total monthly exposure incidence (N events)	max	0.0e+00	NA	NA	NA
Variance in extreme exposure incidence	min	7.1e-05	0.00	NA	NA
Variance in monthly exposure incidence	min	0.0e+00	NA	NA	NA
Time to first extreme exposure occurrence	min	0.0e+00	0.00	NA	NA
Time to first monthly exposure occurrence	min	7.7e-04	NA	NA	NA
Time to last extreme exposure occurrence	min	7.8e-02	0.00	NA	NA
Time to last monthly exposure occurrence	min	1.6e-02	NA	NA	NA
Maximum length of extreme exposure events	max	0.0e+00	0.00	NA	NA
Maximum length of monthly exposure events	max	5.1e-02	NA	NA	NA
Time to extreme exposure peak events	min	0.0e+00	0.12	NA	NA
Time to monthly exposure peak events	min	2.9e-03	NA	NA	NA