Supplementary Material for Improved CMIP6 Projections of Hudson Bay Navigability Using Bias Correction and Route Continuity

Armina Soleymani[a], Alex Crawford[a*], Lawrence Mudryk[b], and Feiyue Wang[a]

[a]Centre for Earth Observation Science, and Department of Environment and Geography, Riddell Faculty of Earth, Environment and Resources, University of Manitoba, Winnipeg, Canada
[b]Climate Research Division, Environment and Climate Change Canada, Toronto, Canada

Correspondence: Alex Crawford (alex.crawford@umanitoba.ca)

Table S1. Summary of CMIP6 models used in this study, including the modeling center, country of origin, realization numbers, and horizontal spatial resolution.

Model	Country	Resolution
ACCESS-CM2	Australia	1.9°×1.3°
BCC-CSM2-MR	China	1.1°×1.1°
CESM2	USA	1.3°×0.9°
CESM2-WACCM	USA	1.3°×0.9°
CMCC-CM2-SR5	Italy	1°×1°
CMCC-ESM2	Italy	1°×1°
CNRM-CM6-1	France	1.4°×1.4°
CNRM-ESM2-1	France	1.4°×1.4°
IPSL-CM6A-LR	France	2.5°×1.3°
MIROC-ES2L	Japan	2.8°×2.8°
MIROC6	Japan	1.4°×1.4°
MPI-ESM1-2-HR	Germany	0.9°×0.9°
MPI-ESM1-2-LR	Germany	1.8°×1.8°
MRI-ESM2-0	Japan	1.1°×1.1°
NorESM2-LM	Norway	2.5°×1.9°
NorESM2-MM	Norway	2°×2°

Section S1. Handling Missing Data

It is necessary to address potential gaps in the input variables—such as OD and CD—which may contain missing values (NaNs). This situation is more common under high-emission scenarios due to extended periods of navigability, and if left uncorrected, it can lead to biased estimates of CD and OD when averaged across years or for multiple models. To preserve the integrity of NP estimates, we implement a correction procedure that adjusts OD and CD symmetrically while maintaining their temporal relationship. First, the "derived" navigable period estimates are inferred from the available percentile-based opening and closing days

$$\mathrm{NP}_{10,\mathrm{derived}} = \mathrm{CD}_{10} - \mathrm{OD}_{90}, \qquad \mathrm{NP}_{50,\mathrm{derived}} = \mathrm{CD}_{50} - \mathrm{OD}_{50}, \qquad \mathrm{NP}_{90,\mathrm{derived}} = \mathrm{CD}_{90} - \mathrm{OD}_{10}. \tag{S1}$$

The difference between the true and derived estimated navigable periods is then calculated as

$$\delta_p = \mathrm{NP}_p - \mathrm{NP}_{p,\mathrm{derived}}, \tag{S2}$$

where $p \in \{10, 50, 90\}$ corresponds to the selected percentile. Assuming the bias is symmetrically distributed between OD and CD, half the bias is then applied symmetrically (in opposite directions) to OD and CD to maintain the true NP

$$\mathrm{OD}_{p,\mathrm{new}} = \mathrm{OD}_p - \frac{\delta_p}{2}, \qquad \mathrm{CD}_{p,\mathrm{new}} = \mathrm{CD}_p + \frac{\delta_p}{2}. \tag{S3}$$

When computing weighted multi-model means, NaN values are excluded from the summations, and the weights are normalized according to the subset of available data. This prevents missing values from artificially biasing the realization mean. The derived NP is compared with the true NP where available, and discrepancies are resolved using the above adjustment scheme (see Supplementary Figures S1 and S2).

Section S2. Experimental Setup

We tuned weighting parameters via grid search over $\sigma_D \in [0.1, 1.0]$ (step 0.01) and $\sigma_S \in [0.1, 1.0]$ (step 0.1). Each pair was tested with leave-one-out cross-validation (LOOCV), withholding one realization as pseudo-truth and predicting it from the rest. This was repeated for all realizations, producing model-specific weights and errors. Predictions were made for 1979-2024, with uncertainty quantified by weighted standard deviation and 95% CIs. Performance was assessed with four metrics—Pearson correlation (ρ), median

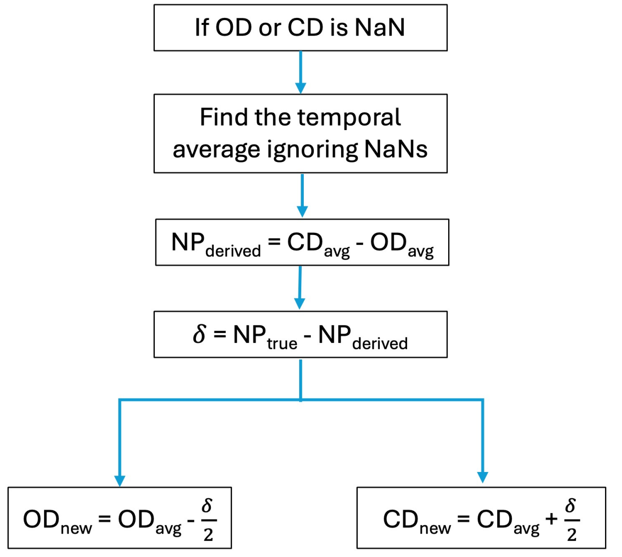


Fig. S1. Procedure for adjusting the average calculation of OD and CD when missing values are present.

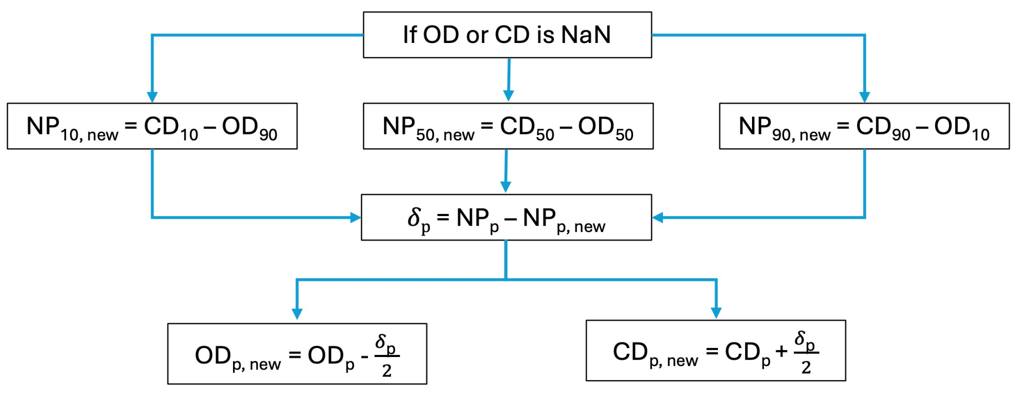


Fig. S2. Procedure for adjusting the percentile calculation of OD and CD when missing values are present.

absolute difference (MAD), median standard error (SE_{median}), and empirical coverage—each min-max normalized. Lower-is-better metrics were inverted. A composite score weighted correlation (0.4) and the other three equally (0.2 each)

$$\mathrm{Score} = 0.4\rho + 0.2 \cdot (1 - \mathrm{MAD}) + 0.2 \cdot (1 - \mathrm{SE}_{\mathrm{median}}) + 0.2\mathrm{CI}_{95\%}. \tag{S4}$$

Model diversity was measured by effective number of models N_{eff}, a metric equivalent to the inverse Simpson index. The Simpson index D (Simpson, 1949), originating from ecology and importance sampling theory, measures dominance. The effective number is calculated as

$$N_{\mathrm{eff}} = \frac{1}{\sum_{i=1}^{N} w_i^2} = \frac{1}{D}, \tag{S5}$$

where w_i represents the normalized weight of realization i, and N is total number of realizations. This measure reflects the number of effectively independent contributors: with equal weights $N_{\mathrm{eff}} \approx N$; if one dominates, $N_{\mathrm{eff}} \to 1$. During optimization, σ_s, σ_d, weights, performance score, and N_{eff} are stored. A threshold requires $N_{\mathrm{eff}} \geq 20\%$ of realizations, preventing dominance by a small subset (e.g., one model). The optimal σ_s and σ_d pair maximizes performance while meeting this constraint, balancing prediction accuracy with structural diversity—both central to robust climate model weighting. Missing values (NaNs) were handled by re-normalizing weights, with NP vs. CD-OD bias corrected as in Section 2.7.1 (see Figures S3 and S4).

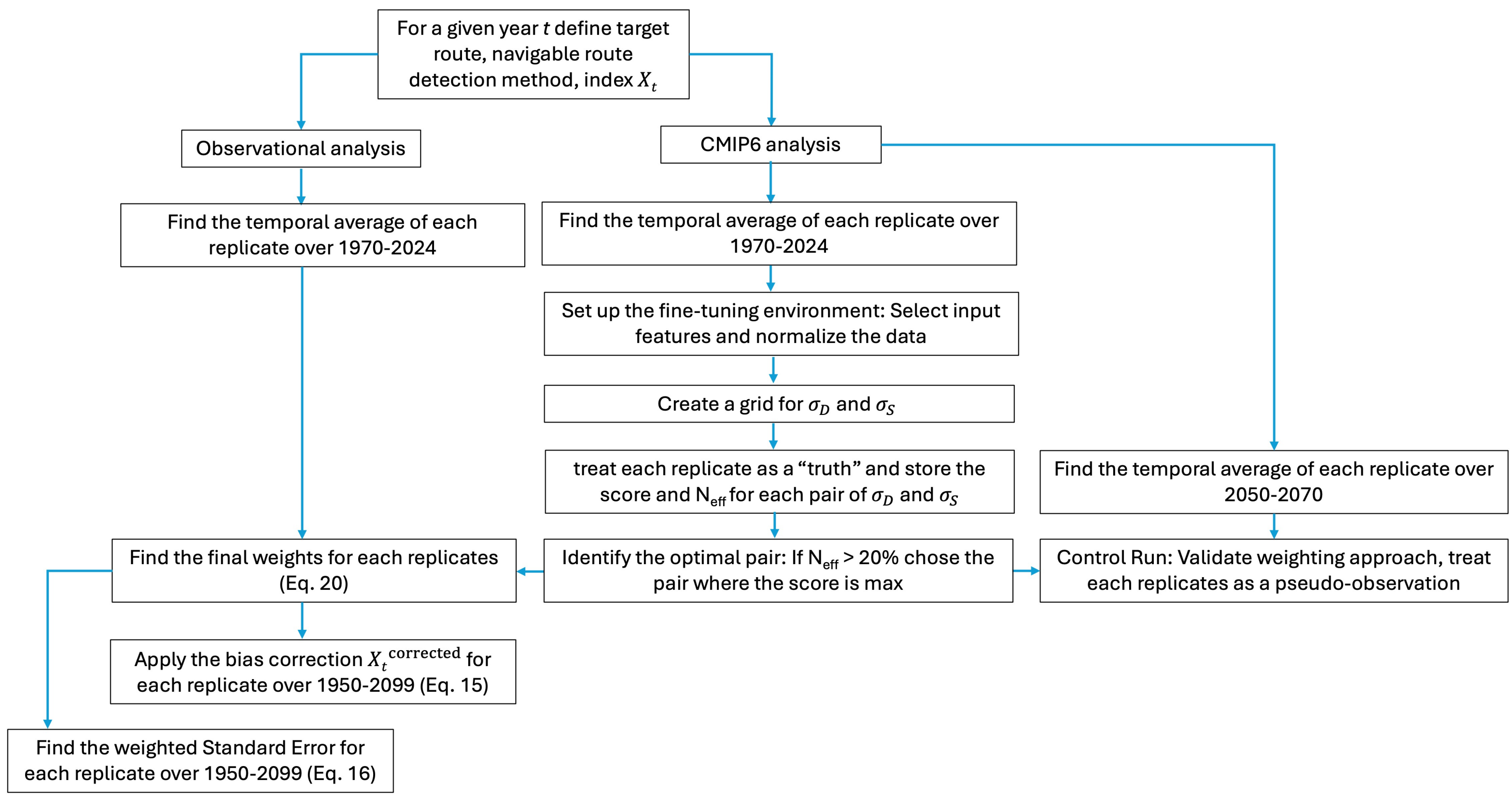


Fig. S3. Workflow of the methodology for weighted average bias correction. This process is repeated for each index, each route, and for each method of detecting route navigability.

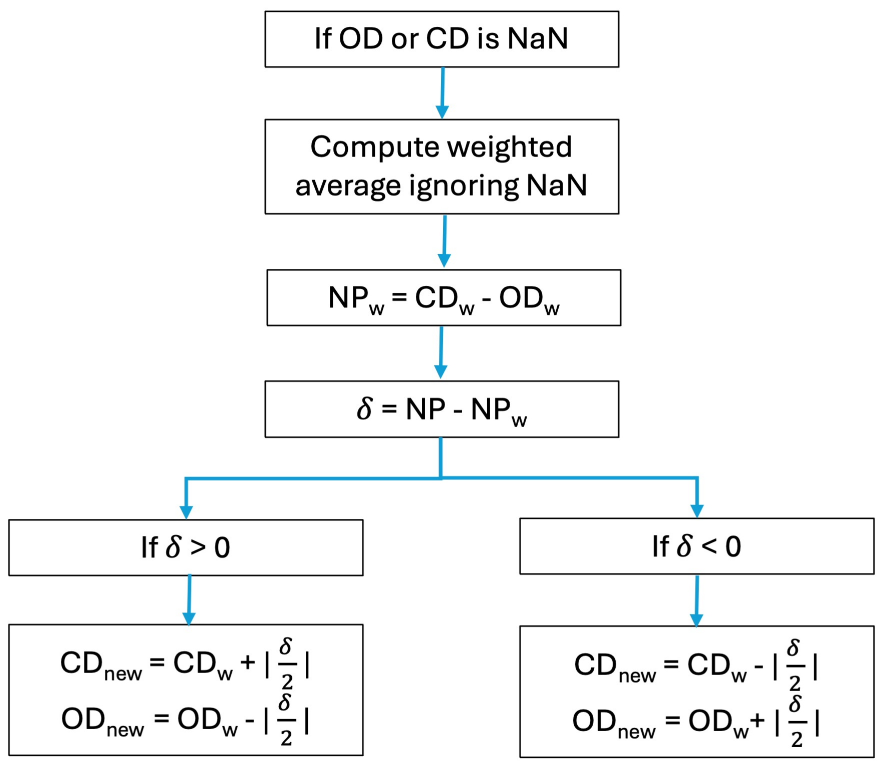
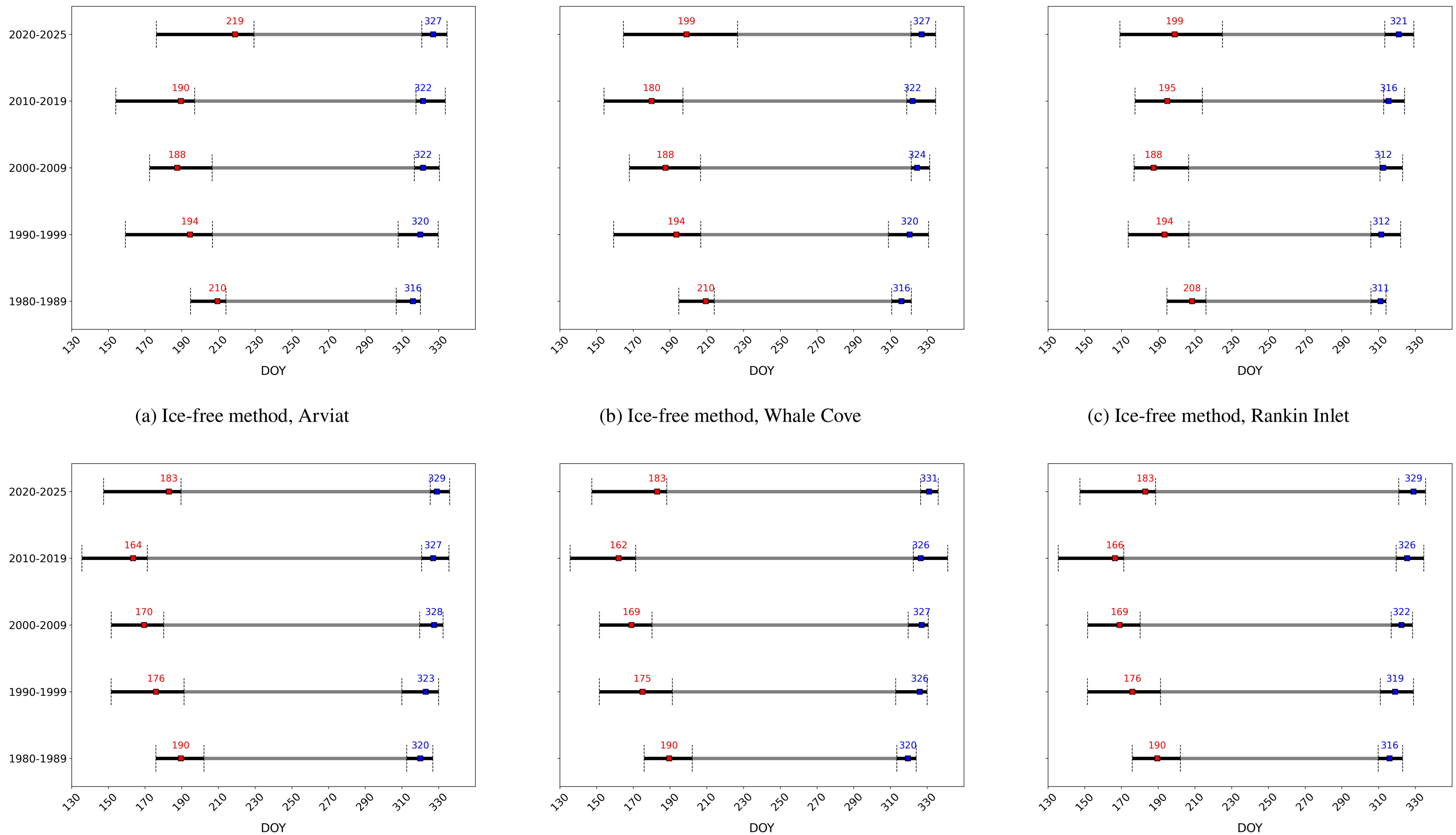


Fig. S4. Procedure for adjusting the weighted averaging calculation of OD and CD when missing values are present.



(a) Ice-free method, Arviat

(b) Ice-free method, Whale Cove

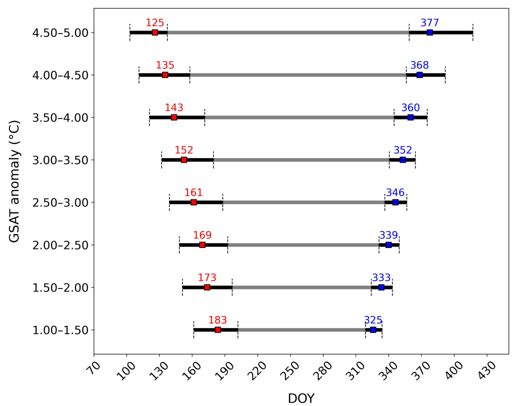
(c) Ice-free method, Rankin Inlet

(d) Type E low-risk method, Arviat

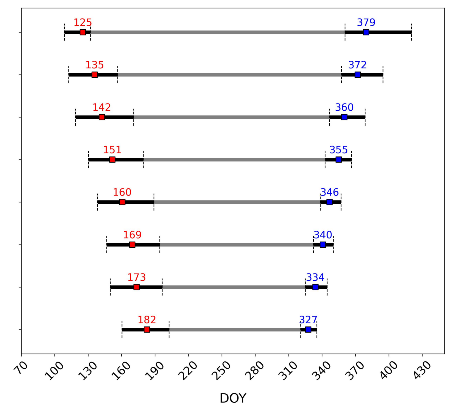
(e) Type E low-risk method, Whale Cove

(f) Type E low-risk method, Rankin Inlet

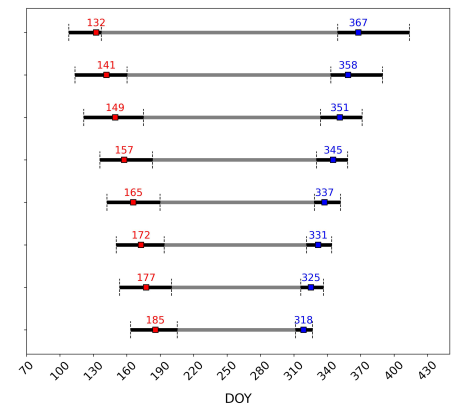
Fig. S5. Historical records of opening and closing days (expressed as day of year, DOY) for maritime routes from the Port of Churchill to: (a) Arviat (ice-free method), (b) Whale Cove (ice-free method), (c) Rankin Inlet (ice-free method), (d) Arviat (Type E low-risk method), (e) Whale Cove (Type E low-risk method), and (f) Rankin Inlet (Type E low-risk method). Red and blue markers represent the median opening day and closing day, respectively, with the interval between them indicating the median navigation period. Error bars denote the 10th-90th percentile range, and the grey bar highlights the period when the route is reliably navigable, except in extreme years.



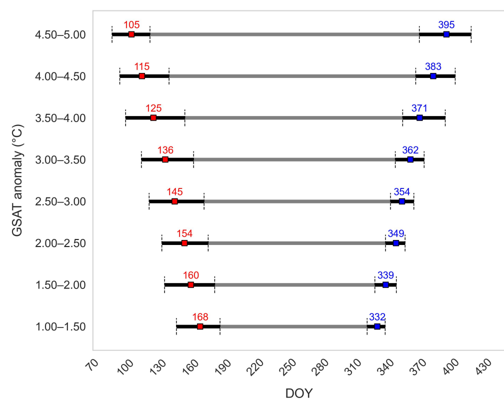
(a) Ice-free method, Arviat



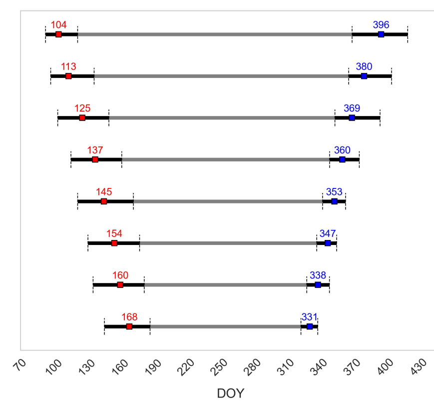
(b) Ice-free method, Whale Cove



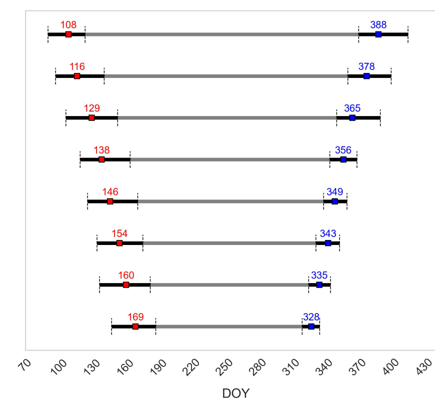
(c) Ice-free method, Rankin Inlet



(d) Type E low-risk method, Arviat

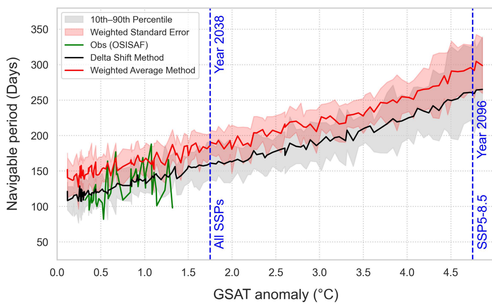


(e) Type E low-risk method, Whale Cove

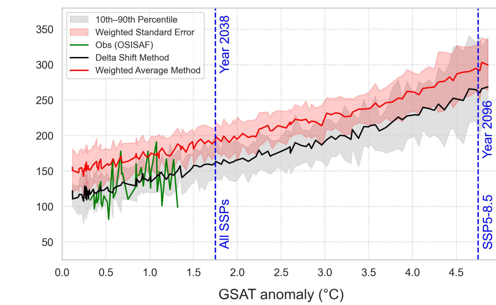


(f) Type E low-risk method, Rankin Inlet

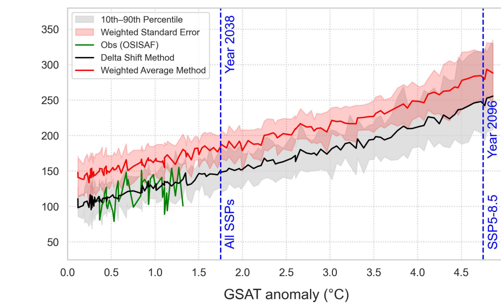
Fig. S6. Projected opening and closing days (day of year, DOY) across global surface air temperature (GSAT) anomaly bins, estimated using the delta-shift bias correction method under SSP5-8.5. GSAT anomalies are relative to the 1850-1900 pre-industrial baseline. Results are shown for routes from the Port of Churchill to: (a) Arviat (ice-free method), (b) Whale Cove (ice-free method), (c) Rankin Inlet (ice-free method), (d) Arviat (Type E low-risk method), (e) Whale Cove (Type E low-risk method), and (f) Rankin Inlet (Type E low-risk method). Red and blue markers represent the median opening day and closing day, respectively, with the interval between them indicating the median navigation period. Error bars denote the 10th-90th percentile range, and the grey bar highlights the period when the route is reliably navigable, except in extreme years.



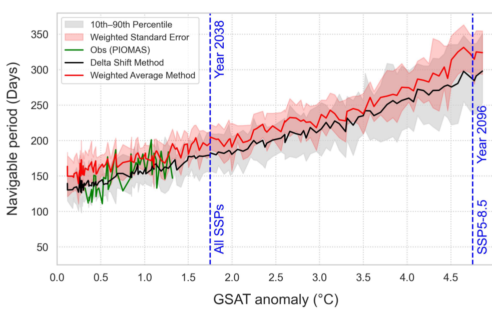
(a) Ice-free method, Arviat



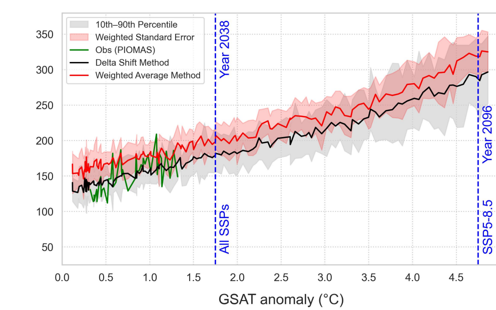
(b) Ice-free method, Whale Cove



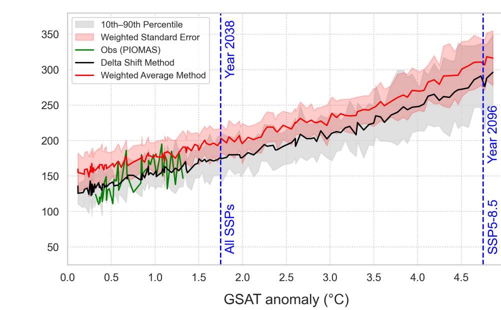
(c) Ice-free method, Rankin Inlet



(d) Type E low-risk method, Arviat



(e) Type E low-risk method, Whale Cove



(f) Type E low-risk method, Rankin Inlet

Fig. S7. Projected navigable period (in days) as a function of global surface air temperature (GSAT) anomaly relative to the 1850-1900 baseline, under SSP5-8.5, for routes from the Port of Churchill to: (a) Arviat (ice-free method), (b) Whale Cove (ice-free method), (c) Rankin Inlet (ice-free method), (d) Arviat (Type E low-risk method), (e) Whale Cove (Type E low-risk method), and (f) Rankin Inlet (Type E low-risk method).

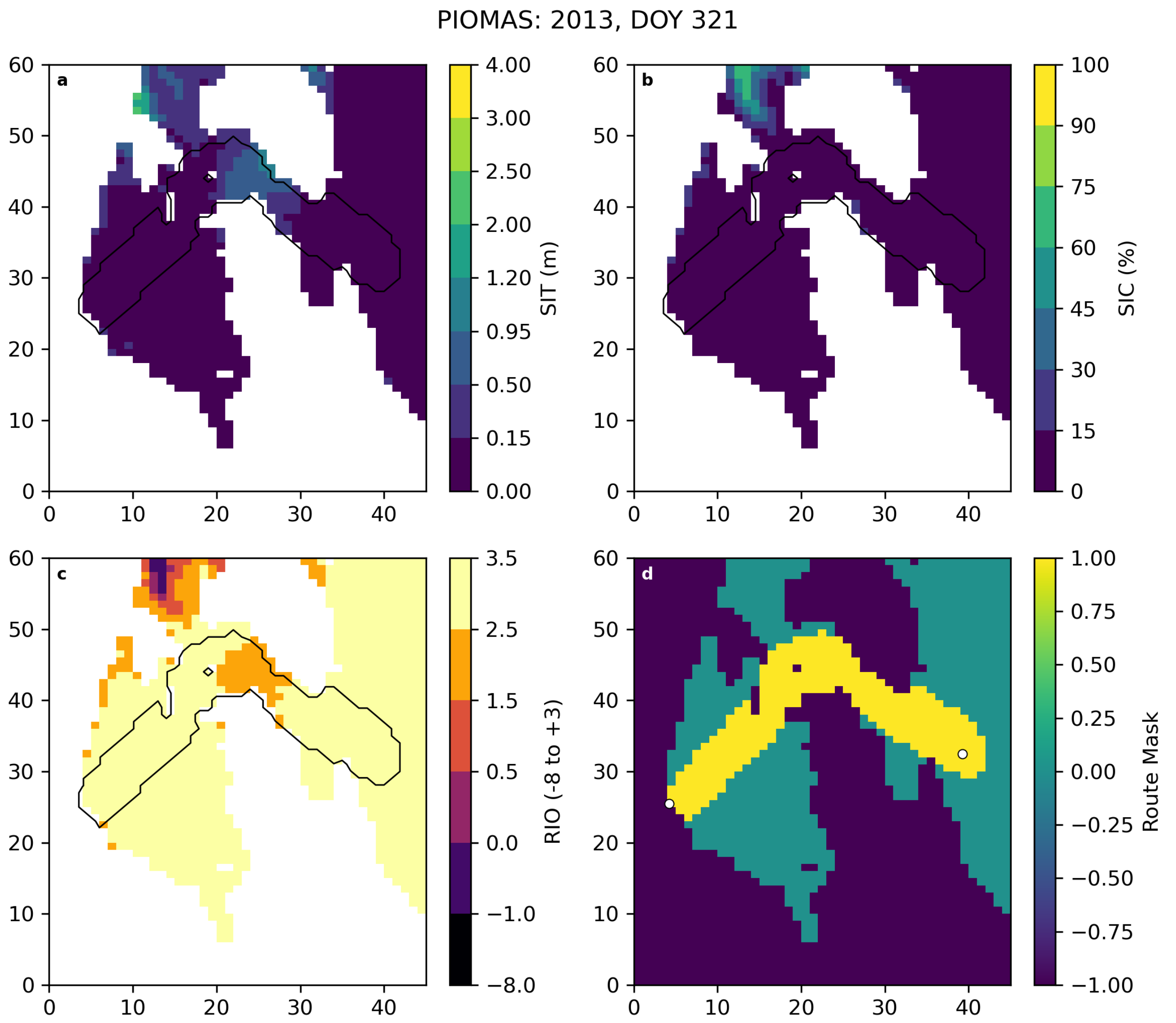


Fig. S8. Example day of PIOMAS (17 November 2013) showing (a) sea ice thickness, (b) sea ice concentration, (c) the RIO using Equation 1, and (d) the study area with start and end nodes (white nodes) of the shipping route and the zone (yellow) used to identify possible pathways. The black outline in (a-c) is a rough match to that shipping zone. On this day, the route would be open based on the ice-free method (SIC < 15%) and open using Equation 1 (low risk method A in Table 2) but closed using Equation 2 (low-risk method B in Table 2).

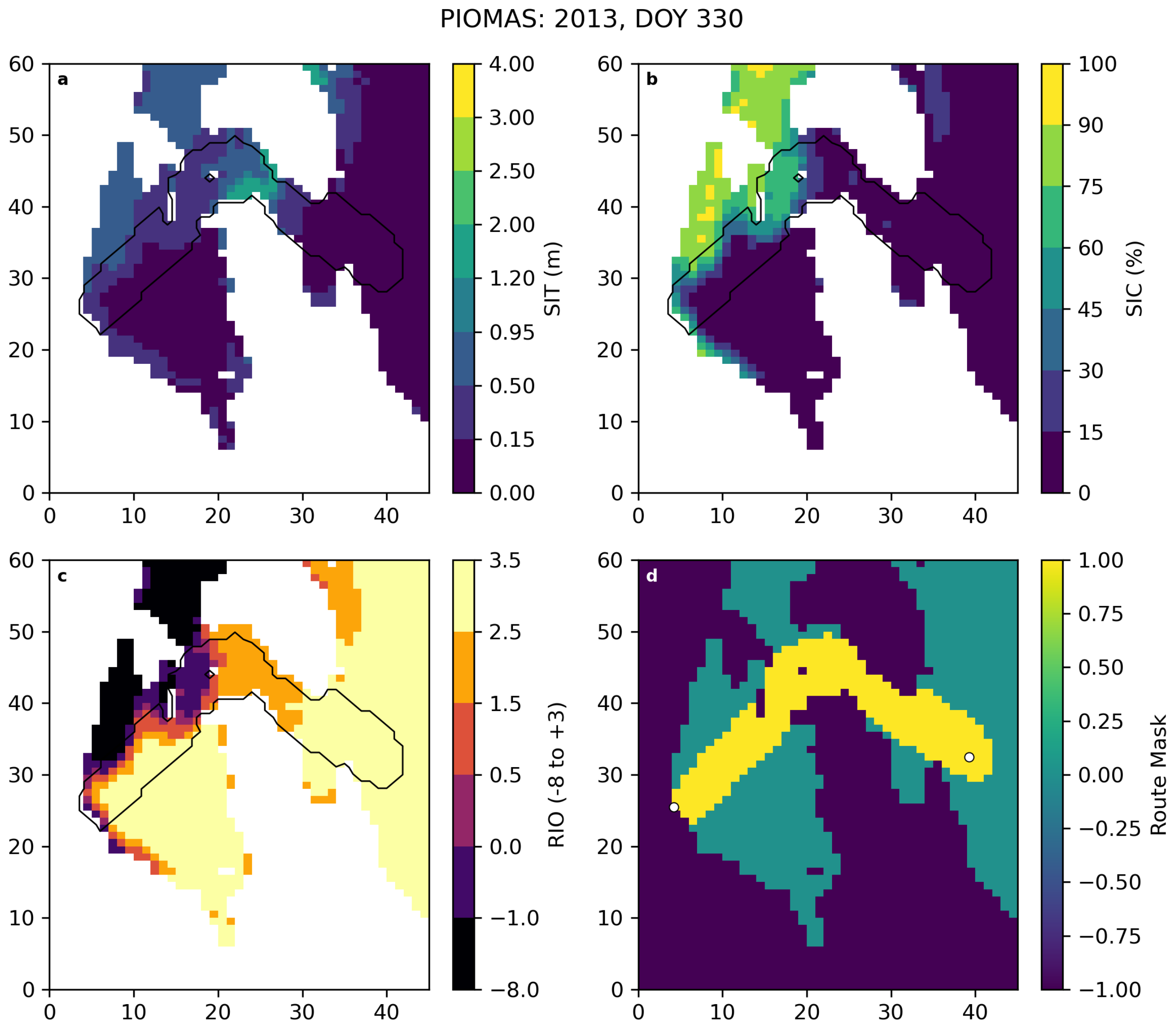


Fig. S9. Example day of PIOMAS (26 November 2013) showing (a) sea ice thickness, (b) sea ice concentration, (c) the RIO using Equation 1, and (d) the study area with start and end nodes (white nodes) of the shipping route and the zone (yellow) used to identify possible pathways. The black outline in (a-c) is a rough match to that shipping zone. This is the closing day if using Equation 2 (low-risk method A in Table 2).

References

Simpson, E. H.: Measurement of diversity, nature, 163, 688–688, 1949.