

SUPPLEMENTARY MATERIAL

Text S1: Determination of time period for present-day climatology

For projecting peak percent cover of *Membranipora membranacea* under present-day sea surface temperature (SST) conditions in the northwest Atlantic Ocean (NWA), we chose to define our own time period because, due to rapid climate warming during the 2010s (NOAA National Centers for Environmental Information 2020), we judged the range of years for “present-day” conditions used by common databases like Bio-ORACLE (2000-2014; Assis et al. 2018) and MARSPEC (2002-2010; Sbrocco & Barber 2013) to no longer be representative of current SST conditions. First, we calculated a long-term baseline temperature for the NWA and compared annual mean temperatures to that baseline. SST data from CoralTemp, which we used to parameterize and project species distribution models (SDMs), are only available from 1985 to present, a period not long enough to establish a long-term baseline. Therefore, for this analysis, we obtained annual SST rasters for the NWA from the NOAA Extended Reconstructed SST (ERSST) v5 dataset (Huang et al. 2017), and calculated a long-term baseline as the mean temperature across all cells between 1900 and 2000. We calculated annual mean temperature across all cells for the range of years in our dataset for percent cover of *Membranipora membranacea* (1987-2020) using the ERSST data and subtracted the long-term baseline temperature from each annual value to obtain annual anomalies. We segregated years into groups based on their temperature anomaly with k-means clustering analysis ($k = 3$, starting iterations = 25; Hartigan & Wong 1979). We found that the most recent cluster of years, with the greatest positive anomaly from 20th century average temperature, was 2010-2020 (Figure S1). Therefore, we selected 2010-2020 as the range over which to calculate present-day percent cover of *M. membranacea*.

Text S2: Future SST Data Processing

To generate estimates of SST 6-max for future time periods (2040-2050 and 2090-2100), we first calculated the difference between mean “present-day” and future SST in the NWA for each cell in the Bio-ORACLE data. We then calculated SST 6-max for each year (as for 2010-2020 SST data in section 2.3) between 2000 and 2014 using the CoralTemp data (to match the Bio-ORACLE “present day” time period) and averaged across years. Lastly, we generated future rasters of SST 6-max by adding the mean temperature difference between present and future scenarios (from Bio-ORACLE) to the mean SST 6-max for 2000-2014 (from CoralTemp). Since the resolution of the Bio-ORACLE data is coarser (0.08° or 9.2 km; Assis et al. 2018) than that of CoralTemp (0.05° or 5 km; Skirving et al. 2020), we used bilinear interpolation to resample the temperature-difference rasters derived from Bio-ORACLE to match the resolution of the CoralTemp data prior to summing the two rasters (Figure S2).

Text S3: Salinity threshold

As *M. membranacea* is thought to be negatively impacted by low salinity (Forbord et al. 2020), and the NWA contains multiple large estuaries, we limited the SDM domain by salinity. We obtained present-day (2000-2014) sea surface salinity (SSS) data from Bio-ORACLE (Assis

et al. 2018) and extracted SSS at all sites in our dataset. When included as a predictor in the SDM, we found no significant relationship between salinity and percent cover, but the sites in our dataset included a narrow range of relatively high salinity (~30–32 PSU). To determine the broader range of salinity inhabited by *M. membranacea* in the NWA, we extracted SSS at monitoring sites with settlement plates colonized by *M. membranacea* in the Gulf of St. Lawrence (Fisheries and Oceans Canada 2006), which has a lower average salinity than the sites in our dataset. The lowest salinity where *M. membranacea* was recorded in the Gulf of St. Lawrence was 21.2 PSU. Since *M. membranacea* exhibits similar phenology and had a similar rate of spread in the Gulf of St. Lawrence compared to other (more saline) areas in the NWA (Gulf of Maine and Nova Scotia; Denley et al. 2019), there is no evidence that *M. membranacea* is negatively impacted by lower salinity levels in the Gulf of St. Lawrence. Therefore, we used 21.2 PSU as the minimum salinity value for projecting SDMs, excluding from the domain the George River Estuary in northern Quebec and much of the St. Lawrence Estuary in southern Quebec.

Text S4: Additional R packages

In addition to those mentioned in the main text, we used several additional R packages to conduct the analyses in this study. We worked with raster data using the *raster* package (Hijmans 2021). We created figures using *ggplot2* (Wickham 2016) and *sf* (Pebesma 2018), and arranged them using *patchwork* (Pedersen 2020). We created tables with *sjplot* (Lüdecke 2020), *kableextra* (Zhu 2021) and *knitr* (Xie 2021). We obtained data from Bio-ORACLE using *sdmpredictors* (Bosch 2021). We used *sp* (Pebesma & Bivand 2005) and *rSDM* (Rodriguez-Sanchez 2021) to assign SST values to sites not on the raster grid. We worked with netcdf files from CoralTemp using *ncdf4* (Pierce 2019). To establish cells in contact with land, we converted the land shapefile to a raster format using *fasterize* (Ross 2020). We conducted k-means clustering analysis using *stats* (R Core Team 2021) and plotted the results using *factoextra* (Kassambara & Mundt 2020). For data management and manipulation, we used *tidyverse* (Wickham et al. 2019), *lubridate* (Grolemund & Wickham 2011), *car* (Fox & Weisberg 2019), *anytime* (Eddelbuettel 2020), and *readxl* (Wickham & Bryan 2019).

Text S5: ODMAP

Overview

Authorship

Contact : c.pratt@dal.ca

Model objective

Model objective: Mapping and interpolation

Target output: Abundance

Focal Taxon

Focal Taxon: *Membranipora membranacea*

Location

Location: Northwest Atlantic Ocean

Scale of Analysis

Spatial extent: -72, -52, 42, 61 (xmin, xmax, ymin, ymax)

Spatial resolution: 5

Temporal extent: - 2010-2020 - 2040-2050 - 2090-2100

Temporal resolution: decadal

Boundary: rectangle

Biodiversity data

Observation type: field survey

Response data type: Abundance

Predictors

Predictor types: climatic, habitat, topographic

Hypotheses

Hypotheses: Cover of *Membranipora* has a positive, exponential relationship with temperature. Cover also depends on the species of kelp acting as its substrate, with *Laminaria digitata* hosting the highest cover, followed by *Saccharina latissima* then *Agarum clathratum*. Cover also has a negative relationship with depth, which is due to decreasing temperature with depth. These hypotheses were all supported by Pratt et al. (2022).

Assumptions

Model assumptions: - All key environmental drivers are included in the model - the species is in equilibrium with its environment - the model extrapolates in a biologically sensible manner; niche conservatism through time - predictors and abundance data were collected without error.

Algorithms

Modelling techniques: generalized linear mixed model (GLMM)

Model complexity: Pratt et al. (2022) conducted a thorough variable selection analysis and determined the predictors used in this study to be the optimal set of predictors for the percent cover of *M. membranacea*

Model averaging: Did not use ensemble modelling because of limited alternatives to GLMM when modelling grouped data. Alternatives that do exist are tree-based (mixed-effects random forest, gaussian process boosting) and not suitable for extrapolation

Workflow

Model workflow: - Fit GLMM according to specifications from Pratt et al. (2022), with minor adjustments - Validate GLMM using either 0.632+ bootstrap (following Potts & Elith 2006) or external validation on test data, depending on analysis - Project GLMM to present and future climate scenarios using rasters of SST and setting depth and kelp substrate as constants

Software

Software: -We used R version 4.0.3. (R Core Team 2021) in the R Studio environment (RStudio Team 2020) for all analyses -key packages: glmmtmb (Brooks et al. 2017), raster (Hijmans 2021)

Code availability: - available from corresponding author upon reasonable request

Data availability: - available from corresponding author upon reasonable request

Data

Biodiversity data

Taxon names: *Membranipora membranacea*

Taxonomic reference system: N/A

Ecological level: populations, species

Data sources: Compiled from literature, unpublished, and original data

Sampling design: Opportunistic (determined by data availability). Nested within sites and years

Sample size: 607 (main dataset); 118 (Norway data)

Clipping: N/A

Scaling: N/A

Cleaning: Percent cover was averaged by site, depth, kelp substrate, and sampling date

Absence data: N/A

Background data: N/A

Errors and biases: Percent cover values will have varying error rate depending on study they originated from

Data partitioning

Training data: No data partitioning was applied

Validation data: 0.632+ bootstrap validation (n = 200 replicates) or external validation used depending on the analysis

Test data: Test data were determined by the needs of the analysis

Predictor variables

Predictor variables: maximum six-month mean sea surface temperature (SST 6-max); depth; kelp substrate

Data sources: daily SST from CoralTemp (NOAA Coral Reef Watch 2020), future SST from Bio-ORACLE (Assis et al. 2018)

Spatial extent: -72, -52, 42, 61 (xmin, xmax, ymin, ymax)

Spatial resolution: CoralTemp: 5km; Bio-ORACLE: 9km

Coordinate reference system: “+proj=longlat +datum=WGS84 +no_defs”

Temporal extent: 1987-2020

Temporal resolution: daily

Data processing: Bio-ORACLE SST rasters downscaled to match resolution of CoralTemp data using bilinear interpolation

Errors and biases: N/A

Dimension reduction: See Pratt et al. (2022)

Transfer data

Data sources: Transferred to Norwegian percent cover data compiled from multiple studies (Førde et al. 2016, Matsson et al. 2019, Forbord et al. 2020)

Spatial extent: 4.7, 19.8, 58.1, 70.1 (xmin, xmax, ymin, ymax)

Spatial resolution: 5km (CoralTemp SST)

Temporal extent: 2014-2017

Temporal resolution: daily SST

Models and scenarios: Transferred NWA model to Norway data to test performance

Data processing: N/A

Quantification of Novelty: N/A

Model

Variable pre-selection

Variable pre-selection: Model averaging-based approach (see Pratt et al. 2022)

Multicollinearity

Multicollinearity: Tested for collinearity using VIF and Pearson correlation

Model settings

GLMM: REML (FALSE), ziformula (~.), formula (percent_cover ~ depth + kelp_substrate + six_month_mean_sst + kelp_substrate*six_month_mean_sst + (1|site) + (1|year)), family (beta_family(link="logit"))

Model settings (extrapolation): N/A

Model estimates

Coefficients: Extracted from glmmTMB output in R

Parameter uncertainty: Standard error of model-averaged parameter estimates (see Pratt et al. 2022)

Variable importance: Model averaged parameter estimates (see Pratt et al. 2022)

Model selection - model averaging - ensembles

Model selection: Information-theoretic approach (see Pratt et al. 2022)

Model averaging: N/A

Model ensembles: N/A

Analysis and Correction of non-independence

Spatial autocorrelation: N/A (autocorrelation not present)

Temporal autocorrelation: N/A (autocorrelation not present)

Nested data: Crossed random intercepts in GLMM to account for nesting within site and year

Threshold selection

Threshold selection: N/A

Assessment

Performance statistics

Performance on training data: Pearson correlation between observed and predicted values, Spearman rank correlation between observed and fitted values, Model calibration (slope and intercept of linear regression between observed and predicted values), RMSE, Average error

Performance on validation data: All statistics estimated via 0.632+ bootstrap for main model validation (section 2.4.2)

Performance on test data: All statistics estimated via test data for comparison to Norwegian percent cover data (section 2.4.4)

Plausibility check

Response shapes: Response plots (see Pratt et al. 2022)

Expert judgement: Map display

Prediction

Prediction output

Prediction unit: peak percent cover of *Membranipora membranacea*

Post-processing: projections clipped to raster cells in contact with land and to areas with salinity greater than or equal to 21.2 (minimum salinity where *M. membranacea* has been recorded in the NWA)

Uncertainty quantification

Algorithmic uncertainty: Standard error

Input data uncertainty: Due to varying methods, uncertainty in input data varies by data point

Parameter uncertainty: Estimates of standard error are low (+/- 10%) and do not have a substantial impact on interpretation of predictions

Scenario uncertainty: for CoralTemp, see Liu et al. (2014) and Maturi et al. (2017); for Bio-ORACLE, see Assis et al. (2018)

Novel environments: N/A

Literature cited

- Assis J, Tyberghein L, Bosch S, Verbruggen H, Serrão EA, De Clerck O (2018) Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography* 27:277–284.
- Bosch S (2021) Sdmpredictors: Species Distribution Modelling Predictor Datasets. R package version 0.2.10. <https://CRAN.R-project.org/package=sdmpredictors>.
- Brooks ME, Kristensen K, Benthem KJ van, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM (2017) GlmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J* 9:378–400.
- Denley D, Metaxas A, Simard N (2019) Ocean temperature does not limit the establishment and rate of secondary spread of an ecologically significant invasive bryozoan in the northwest Atlantic. *Aquat Invasions*:21.
- Eddelbuettel D (2020) Anytime: Anything to “POSIXct” or “Date” Converter. R package version 0.3.9. <https://CRAN.R-project.org/package=anytime>.
- Fisheries and Oceans Canada (2006) Database from the aquatic invasive species monitoring program in Québec. Data published on St Lawrence Global Observatory-SLGO. <https://slgo.ca/> (accessed June 9, 2021)
- Forbord S, Matsson S, Brodahl GE, Bluhm BA, Broch OJ, Handå A, Metaxas A, Skjermo J, Steinhovden KB, Olsen Y (2020) Latitudinal, seasonal and depth-dependent variation in growth, chemical composition and biofouling of cultivated *Saccharina latissima* (Phaeophyceae) along the Norwegian coast. *J Appl Phycol*:1–18.
- Førde H, Forbord S, Handå A, Fossberg J, Arff J, Johnsen G, Reitan KI (2016) Development of bryozoan fouling on cultivated kelp (*Saccharina latissima*) in Norway. *Journal of Applied Phycology* 28:1225–1234.
- Fox J, Weisberg S (2019) *An R Companion to Applied Regression, Third*. Sage, Thousand Oaks CA.
- Grolemund G, Wickham H (2011) Dates and Times Made Easy with lubridate. *J Stat Softw* 40:1–25.
- Hartigan JA, Wong MA (1979) AK-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28:100–108.
- Hijmans RJ (2021) Raster: Geographic Data Analysis and Modeling. R package version 3.4-13. <https://CRAN.R-project.org/package=raster>.
- Huang B, Thorne PW, Banzon VF, Boyer T, Chepurin G, Lawrimore JH, Menne MJ, Smith TM, Vose RS, Zhang H-M (2017) NOAA extended reconstructed sea surface temperature (ERSST), version 5. 1900-2020.
- Kassambara A, Mundt F (2020) Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. <https://CRAN.R-project.org/package=factoextra>.
- Liu G, Heron SF, Eakin CM, Muller-Karger FE, Vega-Rodriguez M, Guild LS, De La Cour JL, Geiger EF, Skirving WJ, Burgess TF (2014) Reef-scale thermal stress monitoring of coral ecosystems: new 5-km global products from NOAA Coral Reef Watch. *Remote Sensing* 6:11579–11606.
- Lüdecke D (2020) SjPlot: Data Visualization for Statistics in Social Science. R package version 2.8.2. <https://CRAN.R-project.org/package=sjPlot>.
- Matsson S, Christie H, Fieler R (2019) Variation in biomass and biofouling of kelp, *Saccharina latissima*, cultivated in the Arctic, Norway. *Aquaculture* 506:445–452.

- Maturi E, Harris A, Mittaz J, Sapper J, Wick G, Zhu X, Dash P, Koner P (2017) A new high-resolution sea surface temperature blended analysis. *Bulletin of the American Meteorological Society* 98:1015–1026.
- NOAA Coral Reef Watch (2020) NOAA Coral Reef Watch Daily Global Satellite Sea Surface Temperature (a.k.a. 'CoralTemp') Version 3.1: Jan. 1, 1986 - Oct. 21, 2019.
- NOAA National Centers for Environmental Information (2020) State of the Climate: Global Climate Report - Annual 2020. NOAA.
- Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. *R J* 10:439–446.
- Pebesma EJ, Bivand RS (2005) Classes and methods for spatial data in R. *R News* 5:9–13.
- Pedersen TL (2020) Patchwork: The Composer of Plots. R package version 1.1.1. <https://CRAN.R-project.org/package=patchwork>.
- Pierce D (2019) Ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files. R package version 1.17. <https://CRAN.R-project.org/package=ncdf4>.
- Potts JM, Elith J (2006) Comparing species abundance models. *Ecological modelling* 199:153–163.
- Pratt CJ, Denley D, Metaxas A (2022) Selection of predictor variables for species distribution models: a case study with an invasive marine bryozoan. *Oecologia* 198:319–336.
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rodriguez-Sanchez F (2021) RSDM: Species distribution and niche modelling in R. R package version 0.3.9. <https://github.com/Pakillo/rSDM>.
- Ross N (2020) Fasterize: Fast Polygon to Raster Conversion. R package version 1.0.3. <https://CRAN.R-project.org/package=fasterize>.
- RStudio Team (2020) RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA.
- Sbrocco EJ, Barber PH (2013) MARSPEC: ocean climate layers for marine spatial ecology: ecological archives E094-086. *Ecology* 94:979–979.
- Skirving W, Marsh B, De La Cour J, Liu G, Harris A, Maturi E, Geiger E, Eakin CM (2020) CoralTemp and the Coral Reef Watch Coral Bleaching Heat Stress Product Suite Version 3.1. *Remote Sensing* 12:3856.
- Wickham H (2016) *Ggplot2: Elegant Graphics for Data Analysis*, 2nd ed. Springer-Verlag, New York.
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the tidyverse. *J Open Source Softw* 4:1686.
- Wickham H, Bryan J (2019) Readxl: Read Excel Files. R package version 1.3.1. <https://CRAN.R-project.org/package=readxl>.
- Xie Y (2021) Knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.36. <https://cran.r-project.org/web/packages/knitr/index.html>.
- Zhu H (2021) KableExtra: Construct Complex Table with “kable” and Pipe Syntax. R package version 1.3.4. <https://CRAN.R-project.org/package=kableExtra>.

Table S1. Model summary table for the species distribution model for peak percent cover of *M. membranacea*, trained on the full dataset from the northwest Atlantic Ocean (section 2.4.1). Predictors were standardized by 2 standard deviations prior to modelling, and a sum-to-zero contrast was set for kelp substrate. Due to sum-to-zero contrast for kelp substrate, statistics for kelp substrate and kelp:sea surface temperature (SST) use the intercept and the main effect for SST as reference categories, respectively. AC = *A. clathratum*, LD = *L. digitata*, SL = *S. latissima*. Significant p-values in **bold** ($\alpha = 0.05$). σ^2 is the residual variance and τ is the between-group variance for random intercepts.

Predictors	Estimates	Std. error	z value	p-value
Conditional Model				
(Intercept)	-2.36	0.15	-16.21	<0.001
Depth	-0.36	0.10	-3.70	<0.001
Kelp (AC)	-0.25	0.10	-2.53	0.011
Kelp (LD)	0.38	0.08	4.53	<0.001
Kelp (SL)	-0.13	0.07	-1.85	0.064
Six-Month Mean SST	1.79	0.14	13.18	<0.001
Kelp (AC): Six-Month Mean SST	-0.59	0.20	-2.88	0.004
Kelp (LD): Six-Month Mean SST	0.61	0.17	3.56	<0.001
Kelp (SL): Six-Month Mean SST	-0.02	0.14	-0.153	0.879
Zero-Inflation Model				
(Intercept)	-3.66	0.78	-4.67	<0.001
Depth	-0.22	0.44	-0.51	0.611
Kelp (AC)	2.72	0.60	4.51	<0.001

Predictors	Estimates	Std. error	z value	p-value
Kelp (LD)	-3.11	1.05	-2.95	0.003
Kelp (SL)	0.39	0.58	0.68	0.499
Six-Month Mean SST	-7.40	1.14	-6.48	<0.001
Kelp (AC): Six-Month Mean SST	3.29	1.13	2.90	0.004
Kelp (LD): Six-Month Mean SST	-5.52	1.95	-2.83	0.005
Kelp (SL): Six-Month Mean SST	2.23	1.08	2.08	0.038
Random Effects				
σ^2	0.45			
τ_{00} site	0.23			
τ_{00} year	0.06			

Table S2. Model summary table for the species distribution model for peak percent cover of *M. membranacea*, trained on the Norway dataset (section 2.4.4). Predictors were standardized by 2 standard deviations prior to modelling. Significant p-values in **bold** ($\alpha = 0.05$). σ^2 is the residual variance and τ is the between-group variance for random intercepts.

Predictors	Estimates	Std. error	z value	p-value
Count Model				
(Intercept)	-1.47	0.16	-9.32	<0.001
Depth	0.73	0.23	3.12	0.002
Six-Month Mean SST	1.62	0.31	5.21	<0.001
Zero-Inflated Model				
(Intercept)	-1.25	0.30	-4.24	<0.001
Depth	0.68	0.49	1.38	0.166
Six-Month Mean SST	-4.61	0.91	-5.07	<0.001
Random Effects				
σ^2	0.49			
τ_{00} site	0.00			
τ_{00} Year	0.00			

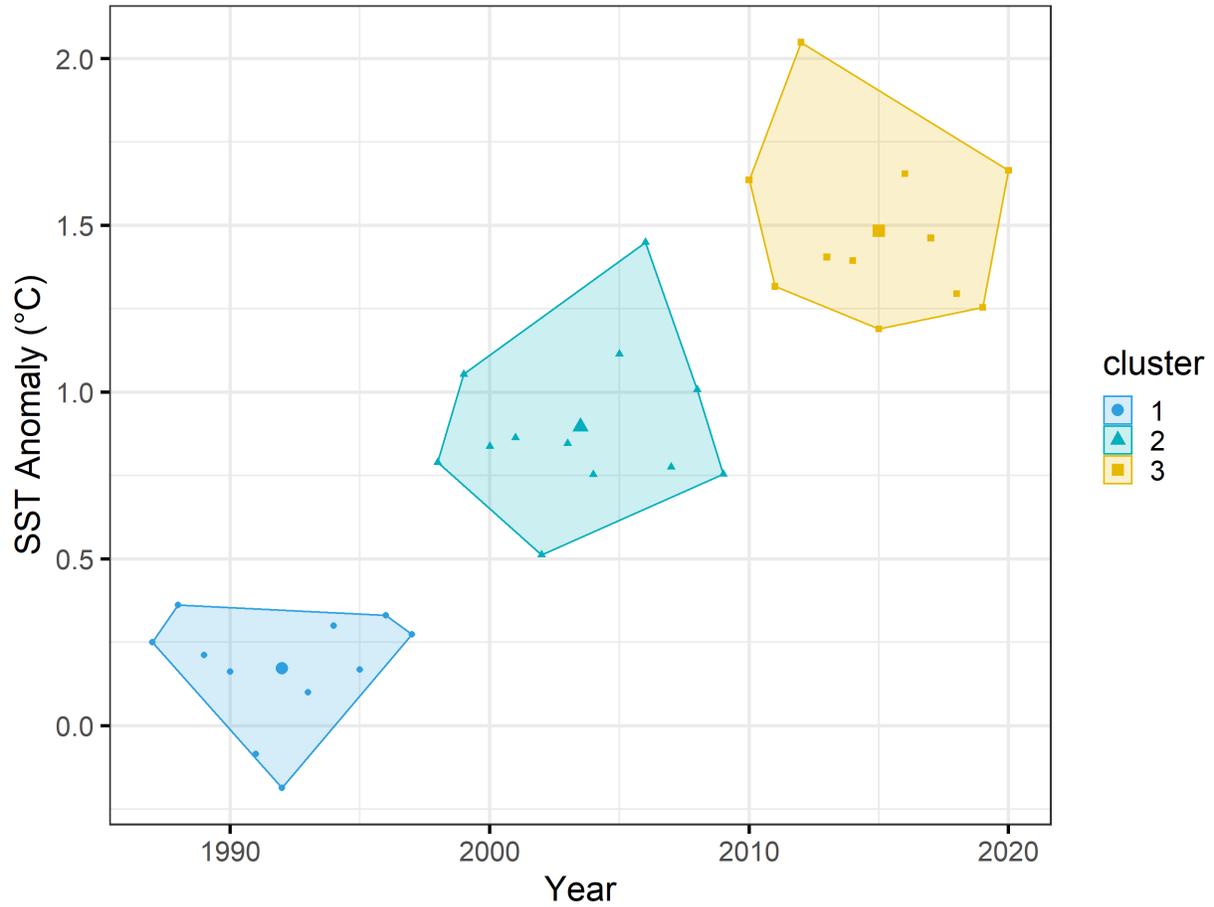


Figure S1. Results of k-means clustering of annual sea surface temperature (SST) anomalies from mean SST over 1900-2000. Clusters 1-3 are groups of years with similar values of SST anomaly, where small points represent annual mean SST values and large points represent the cluster mean.

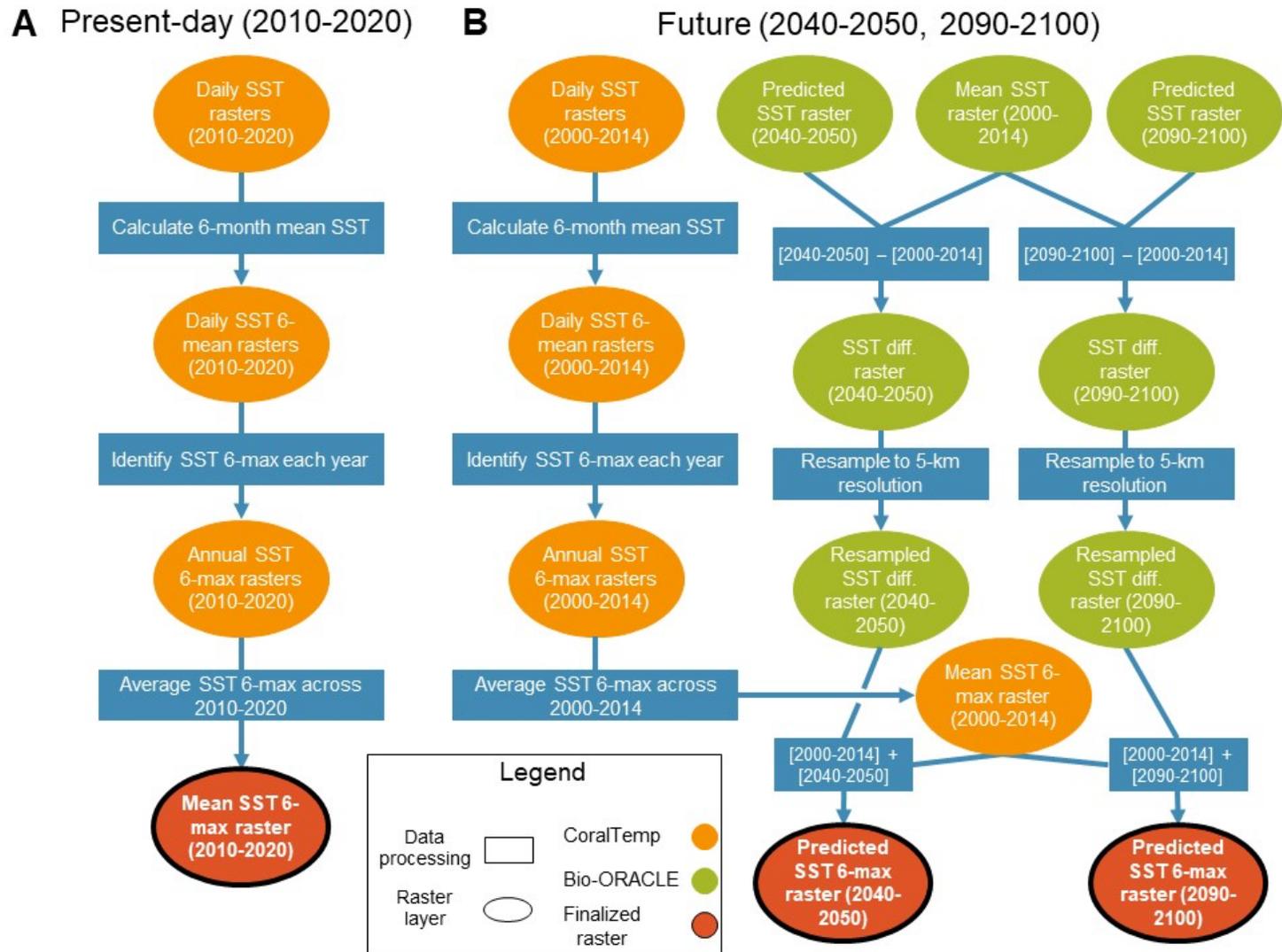


Figure S2. Steps for processing sea surface temperature (SST) data for projecting species distribution models (SDMs) under present-day (A) and future (B) SST conditions. For each future time period (2040-2050 and 2090-2100), the procedure in B was conducted under both Representative Concentration Pathways 4.5 and 8.5.

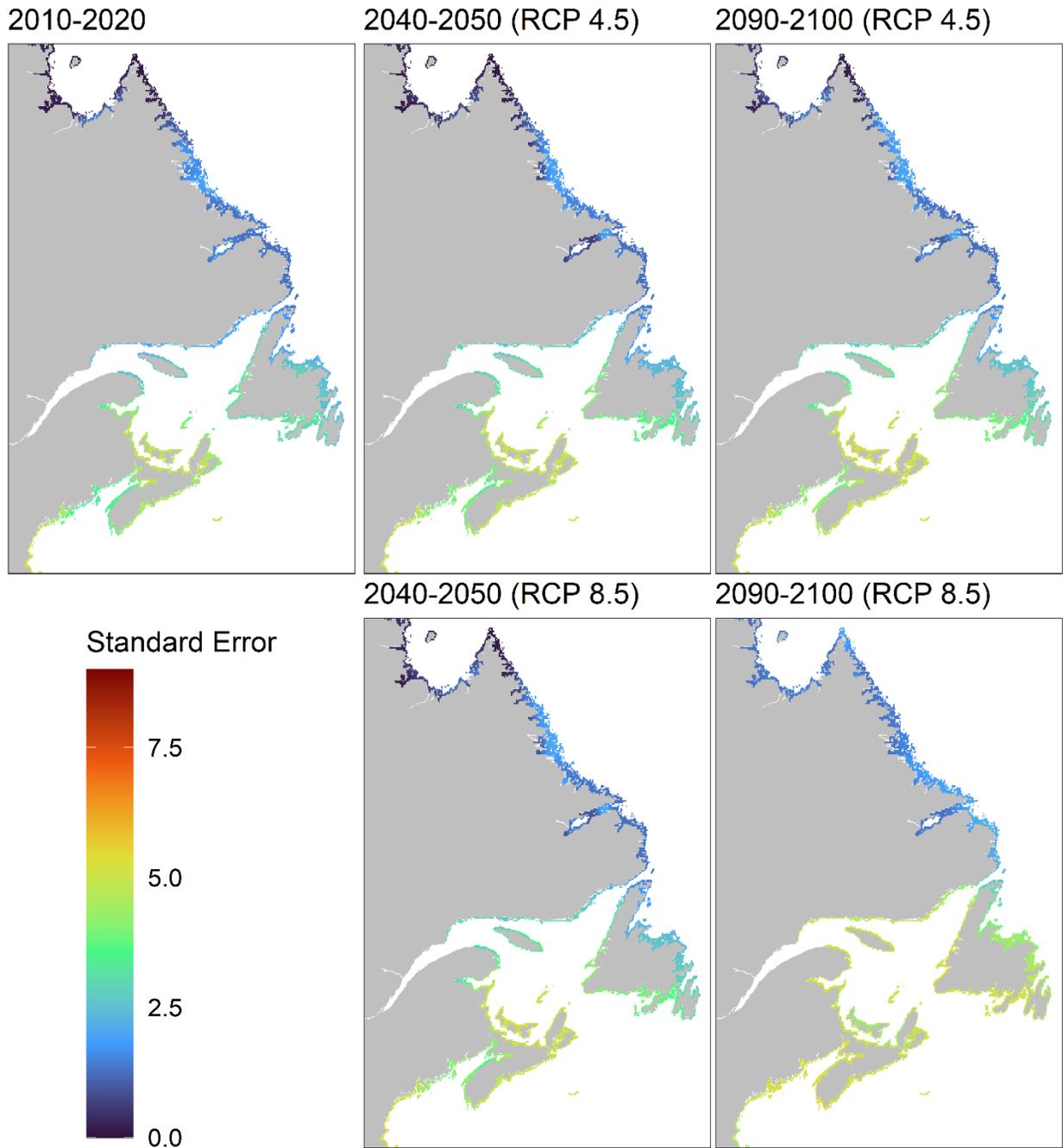


Figure S3. Model standard error for species distribution model projections of peak percent cover of *M. membranacea* on *L. digitata* for present-day (2010-2020) and future sea surface temperature (SST) scenarios. RCP is the Representative Concentration Pathway under which SST has been predicted for each future time period.

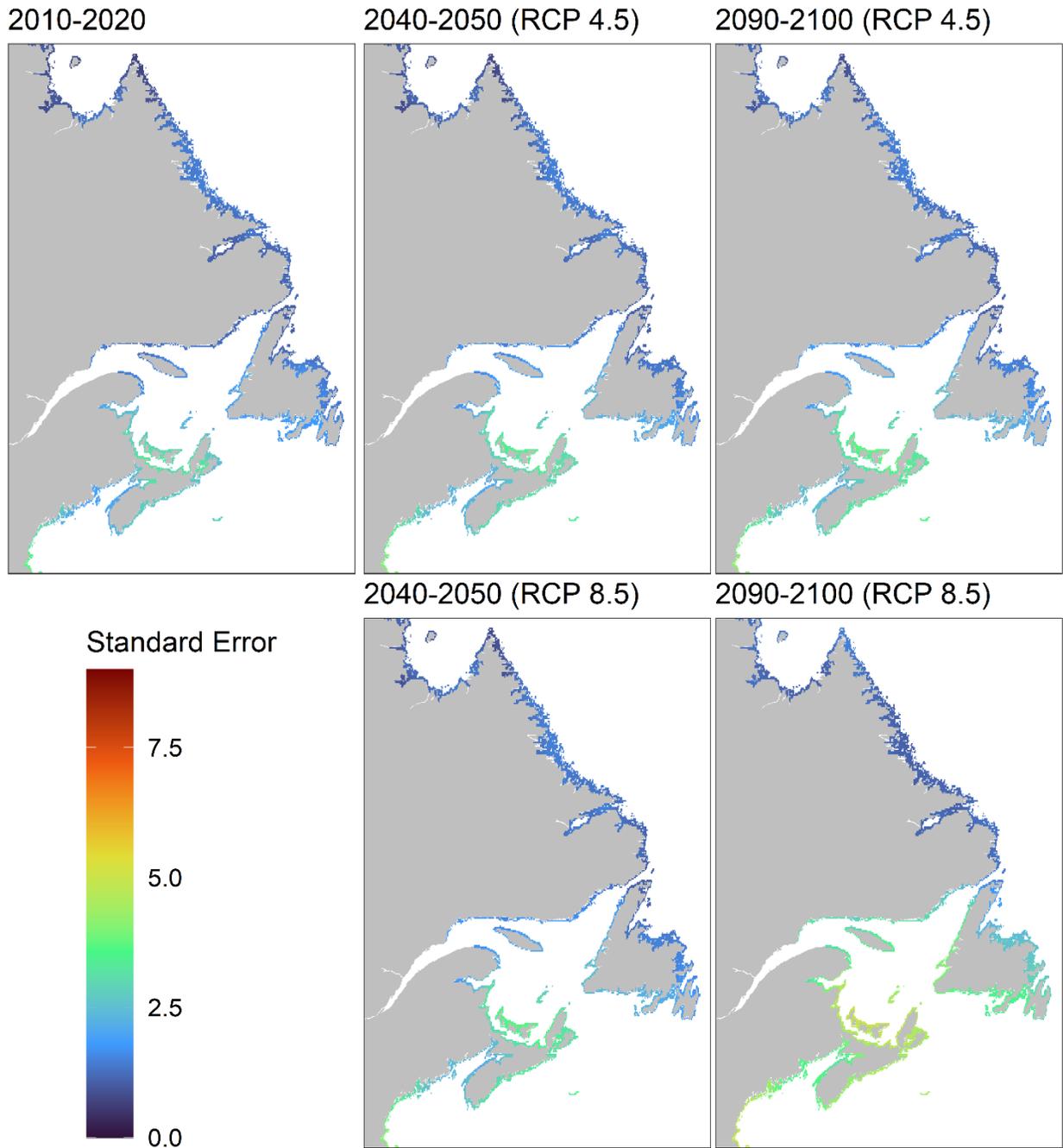


Figure S4. Model standard error for species distribution model projections of peak percent cover of *M. membranacea* on *S. latissima* for present-day (2010-2020) and future sea surface temperature (SST) scenarios. RCP is the Representative Concentration Pathway under which SST has been predicted for each future time period.

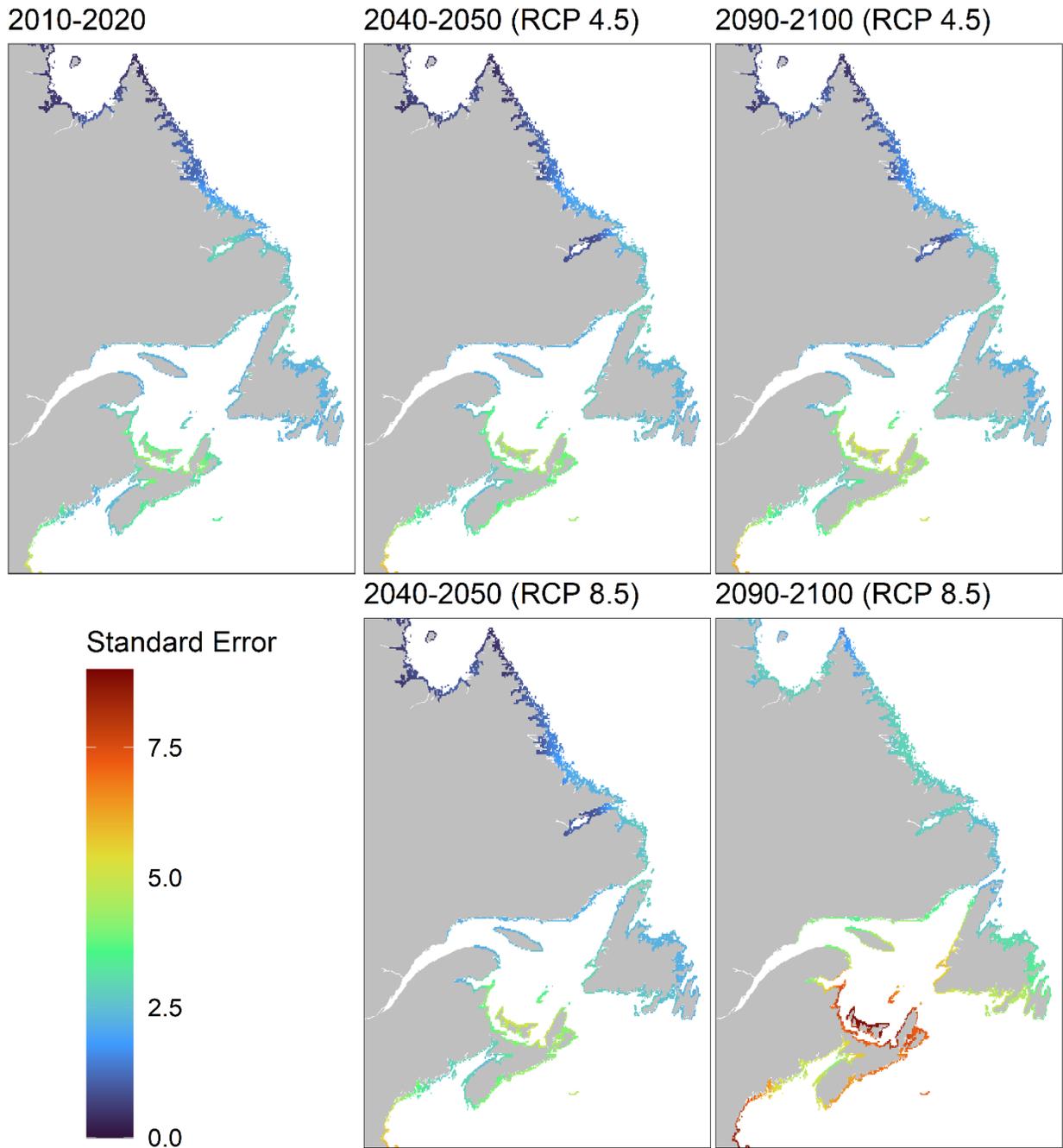


Figure S5. Model standard error for species distribution model projections of peak percent cover of *M. membranacea* on *A. clathratum* for present-day (2010-2020) and future sea surface temperature (SST) scenarios. RCP is the Representative Concentration Pathway under which SST has been predicted for each future time period.

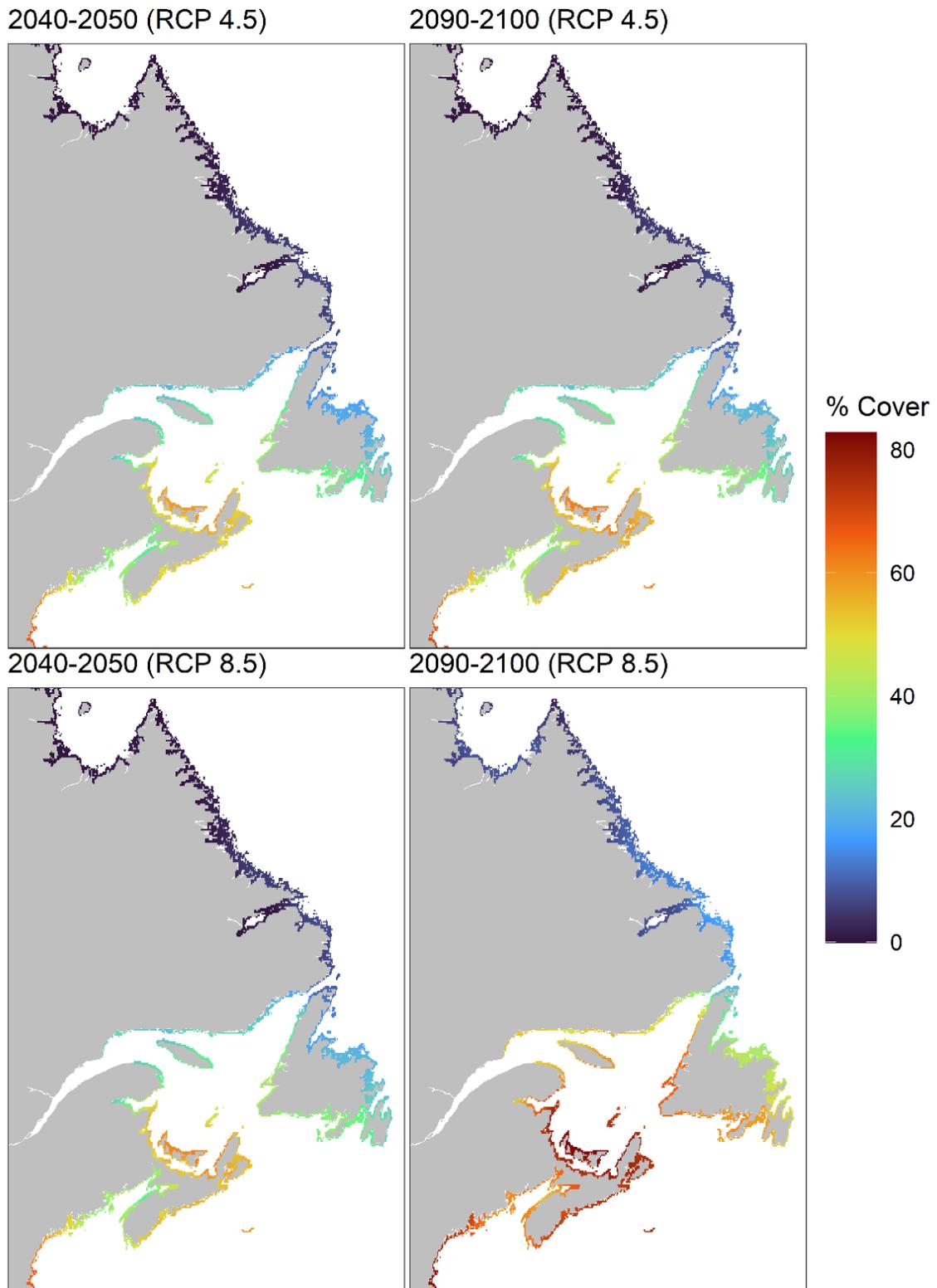


Figure S6. Species distribution model projections of peak percent cover of *M. membranacea* on *L. digitata* for future sea surface temperature (SST) scenarios. RCP is the Representative Concentration Pathway under which SST has been predicted for each future time period.

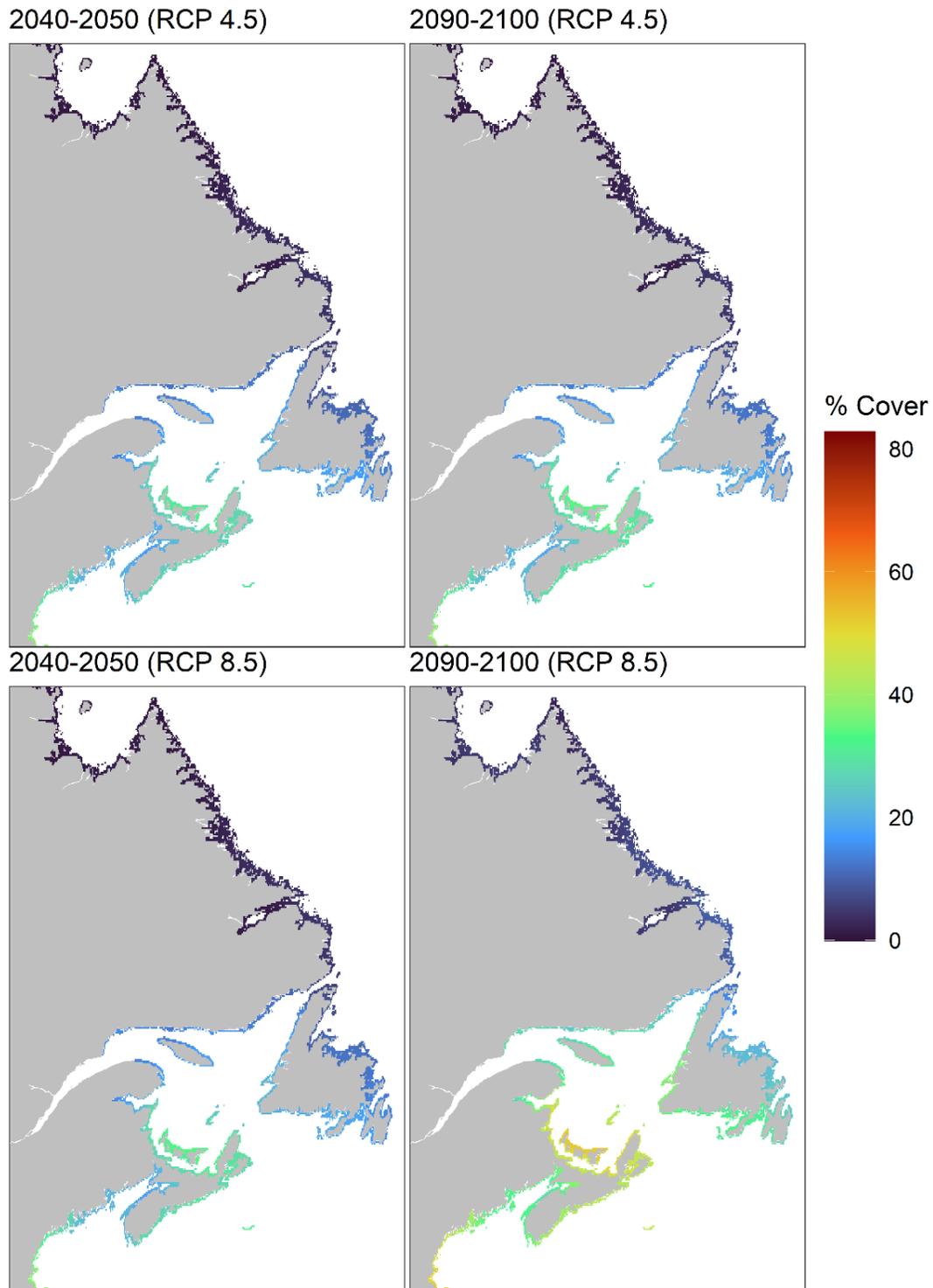


Figure S7. Species distribution model projections of peak percent cover of *M. membranacea* on *S. latissima* for present-day (2010-2020) and future sea surface temperature (SST) scenarios. RCP is the Representative Concentration Pathway under which SST has been predicted for each future time period.

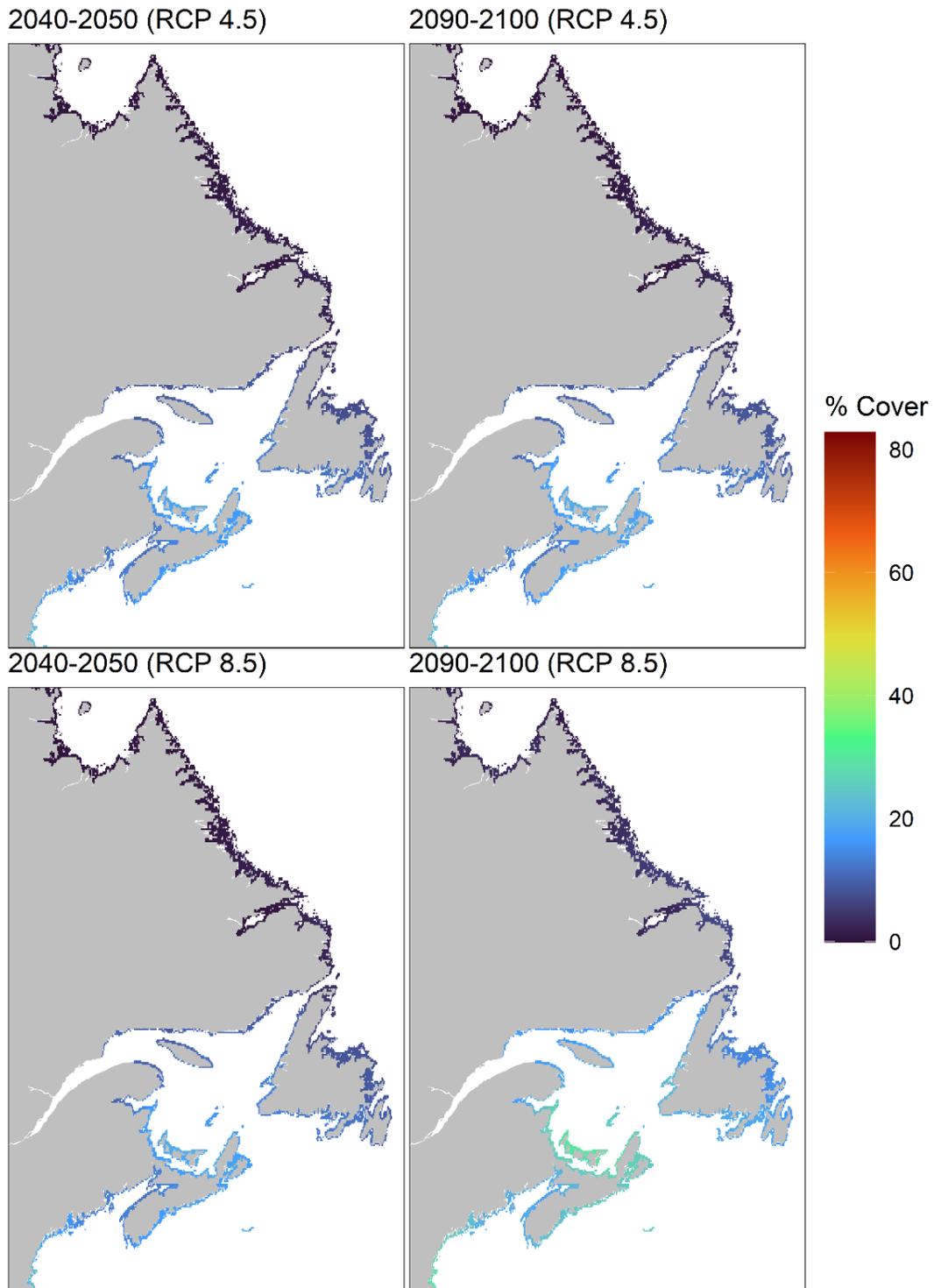


Figure S8. Species distribution model projections of peak percent cover of *M. membranacea* on *A. clathratum* for present-day (2010-2020) and future sea surface temperature (SST) scenarios. RCP is the Representative Concentration Pathway under which SST has been predicted for each future time period.