This Supplement includes:

Section S1: Supplemental figures to the main paper (Figs. S1–S4).
Section S2: Supplemental text (Text S1) with associated Tables and Figures (Tables S1 & S2, Figs. S5–S7)

## Section S1.

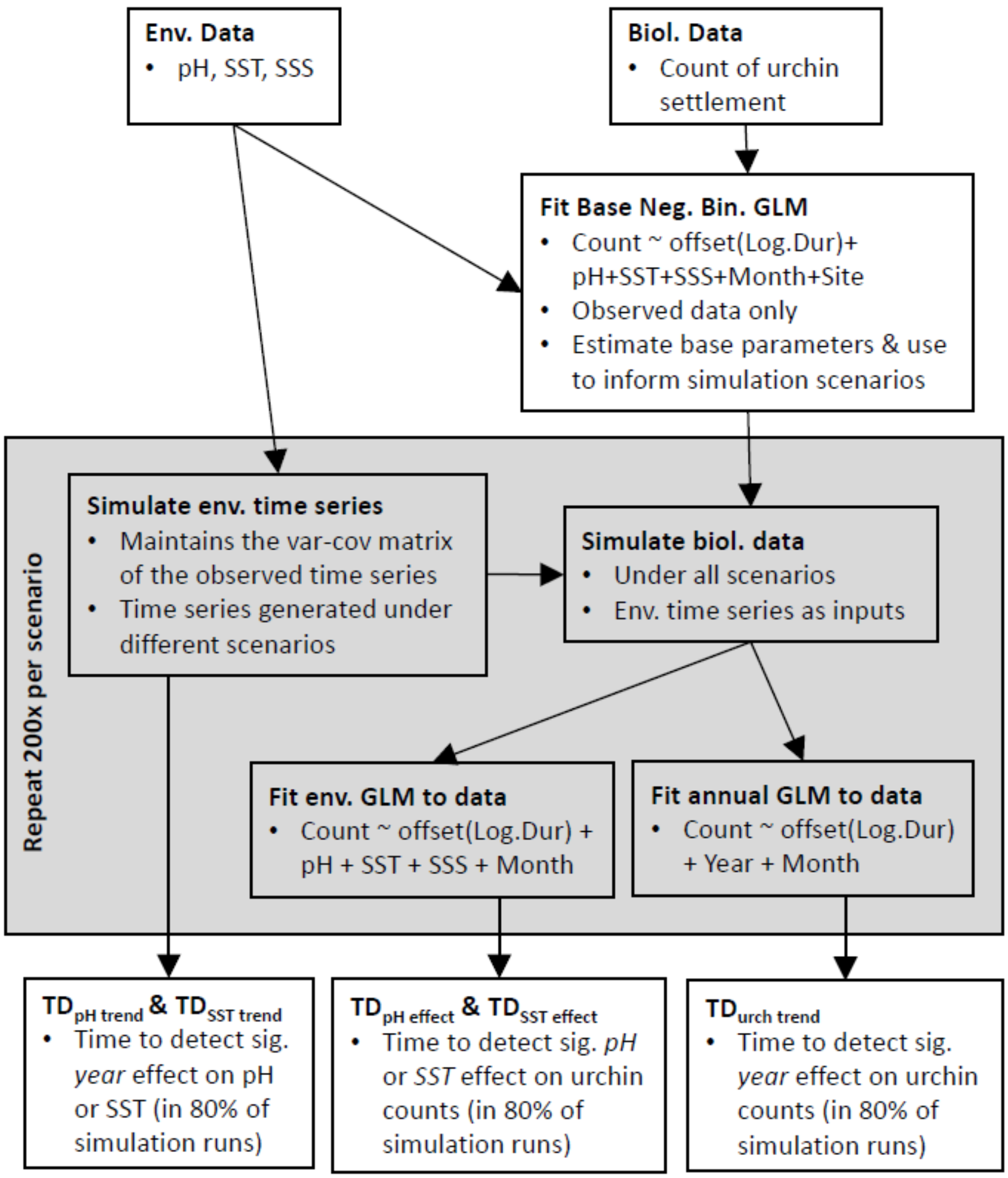


Fig. S1. Conceptual flow diagram for the modeling and simulation process to calculate the time to detect (TD) annual trends in pH and sea surface temperature (SST), TD for pH and SST effects on urchin counts, and TD annual trends in urchin counts in Generalized Linear Models (GLM).

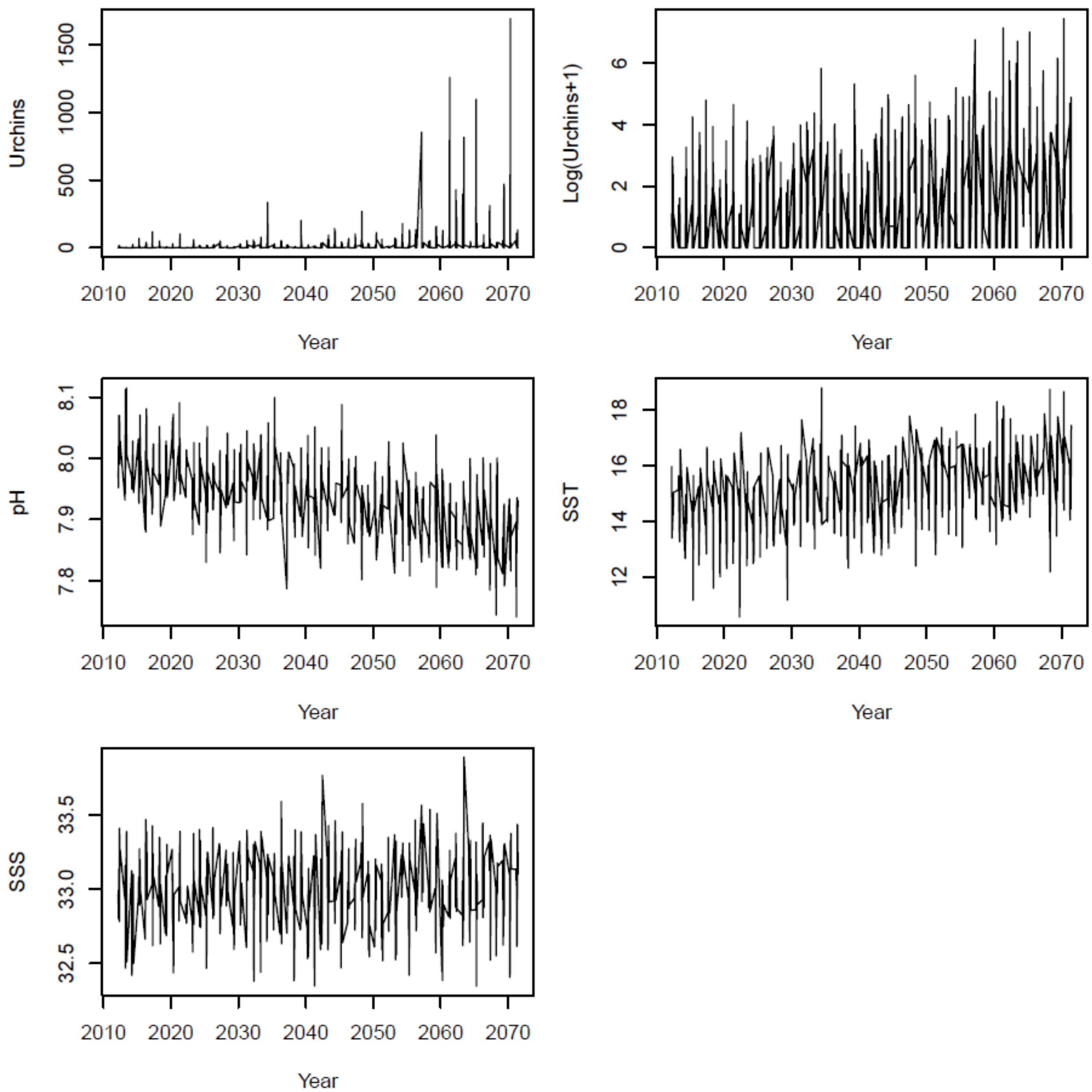


Fig. S2. An example set of simulated time series for urchin settlement count, pH, sea surface temperature (SST in °C), and sea surface salinity (SSS in psu). Data were generated for 60 years, for the following scenarios: base model's coefficients for pH, SST, and dispersion (i.e., relative effect size = 1), intermediate scenarios for long-term trends in pH (-0.1 per 50 yr, or -0.002/yr) and SST (1.5°C per 50 yr, or 0.03°C/yr), and the baseline pH-SST correlation (r=0.34).

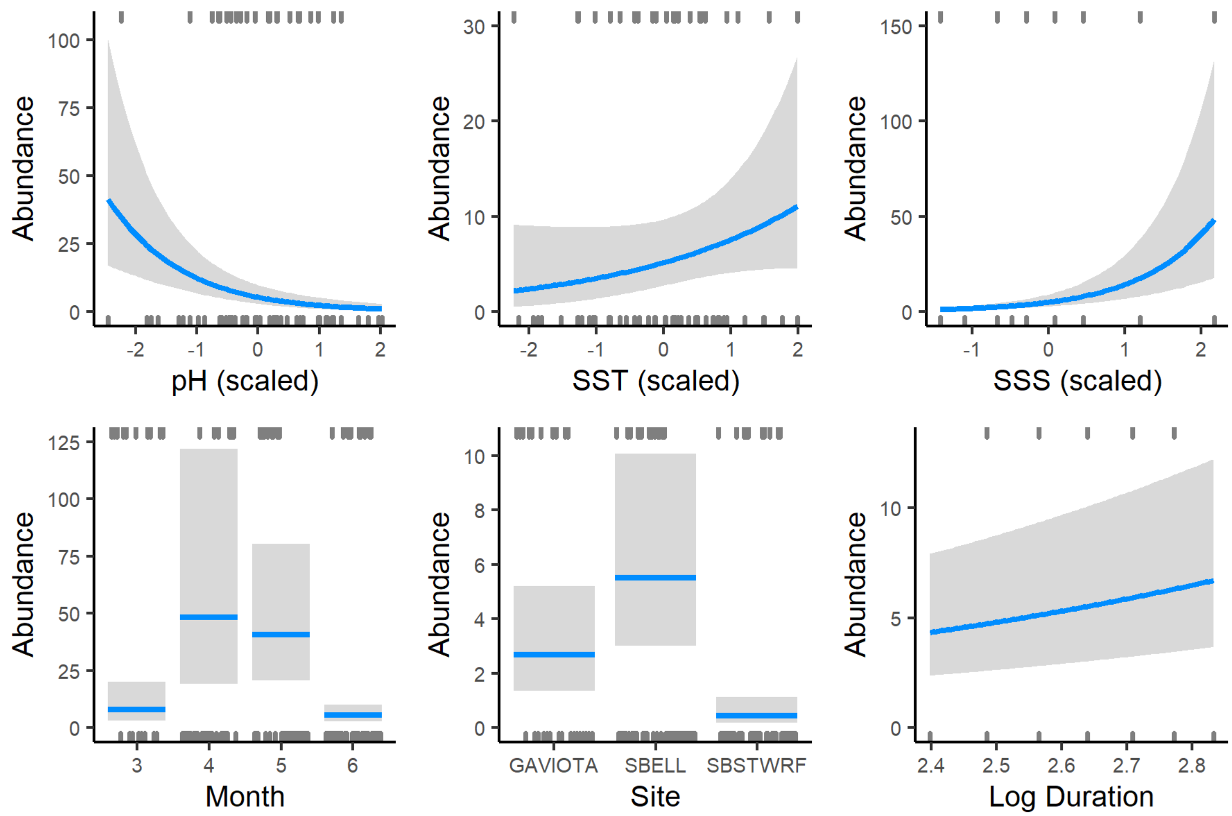


Fig. S3. Response plot on the scale of the response for the negative binomial GLM of urchin abundance as a function of pH, SSS, SST, Month, and Site. Environmental covariates were scaled to mean=0 and SD=1 prior to fitting. Each panel represents predictions (blue lines) with 95% confidence intervals (shading), calculated while holding the other variables at their mean (or median) value. The rugs represent the location of the positive and negative residuals in each panel.

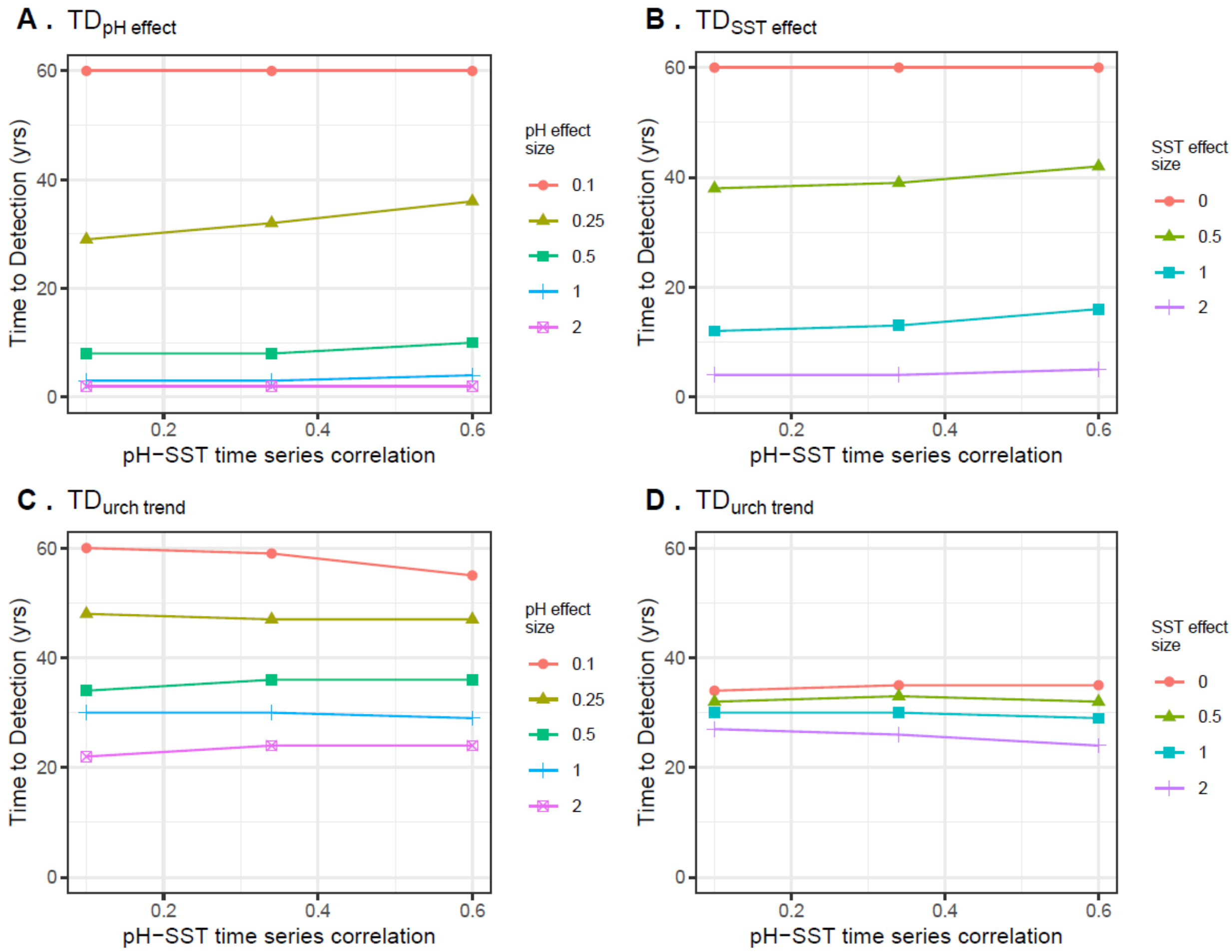


Fig. S4. Time to detect a significant pH effect (A), SST effect (B), and annual trends in urchin counts (C, D) in statistical models fit to simulated data under different scenarios. Panels focus on the effect of the long-term pH-SST time series correlation (X-axis), and either the pH or SST effect sizes (colors) while holding the other scenario factors constant at their baseline values. The maximum number of years to detection was 60, representing values ≥ 60.

# Section S2.

## Text S1. Time of Emergence for environmental pH and SST time-series

Time of emergence (ToE) is an alternative approach to assess how many years it would take for a long-term linear trend in the environmental time series to emerge from the background noise of natural variability (Keller et al. 2014, Carter et al. 2019, Sutton et al. 2022). ToE was calculated for comparison with our estimates of $TD_{ph\ trend}$ and $TD_{SST\ trend}$.

ToE was calculated as:

$$ToE = \frac{2N}{|S|} \quad \text{[Eq. 1]}$$

where ToE is measured in years, N is a measure of noise or variability, and S is the signal or linear trend (per year) in a time-series. N can be composed of various sources of variability or uncertainty, including measurement uncertainty, temporal variability (at any number of scales such as daily, monthly, annual), or spatial variability (Carter et al. 2019). The use of 2N acts as an approximate 95% confidence criterion (instead of the more precise 1.96*N where N is measured as a standard deviation). N for each time-series was calculated as:

$$N = \sqrt{\bar{V}_{2wk}^2 + RMSE^2} \quad \text{[Eq. 2]}$$

such that N was composed of two primary sources of uncertainty: (1) the average variability (i.e., standard deviation, SD) of the biweekly measurements of the data ($\bar{V}_{2wk}$), and (2) the uncertainty in the linear regression calculated as the root mean square error (RMSE) of the regression. $\bar{V}_{2wk}$ was calculated for the empirical pH and SST time-series to capture the average variability within the biweekly estimates of the environmental data, and it was calculated as an average of the standard deviations (with unequal sample sizes):

$$\bar{V}_{2wk} = \sqrt{\frac{\sum_{i=1}^{k}(n_i - 1)V_i^2}{\bar{n}(k-1)}} \quad \text{[Eq. 3]}$$

where $V_i^2$ is the square of the SD of the measurements for biweekly time period *i*, $n_i$ is the number of individual measurements collected within biweekly period *i* (typically n=1008 data points, measured every 20 minutes for two weeks), *k* is the total number of biweekly periods in the data set, and $\bar{n}$ is the average number of individual measurements across all *k* biweekly periods.

ToE was also estimated for the pH and SST time series after detrending the time series for month effects, following Carter et al. 2019. Detrended pH and SST time series were constructed by calculating monthly anomalies from the overall time series mean, and subtracting the monthly anomaly from each biweekly measure of pH and SST. However, to account for the interannual variability in monthly means, an interannual variability term ($\bar{V}_I$) was added in quadrature to Eq. 2. $\bar{V}_I$ was calculated as the average of standard deviations for each month across years using Eq. 3 (with the summation from *m*=1 to 12 instead of *i*=1 to *k*).

ToE for pH ($ToE_{ph}$) and SST ($ToE_{SST}$) were estimated using both the empirical time series (for each of the three sites separately) and the simulated time series. For the simulated data, the median ToE across 200 simulated datasets was used, and this was done for short time series (6 years; similar length to the empirical time series) and longer time series (20 years) because ToE estimates vary based on the length of time series. ToE was calculated using both the raw and monthly-detrended time series. ToE was not calculated for the urchin settlement time series because it is inappropriate for discrete, non-normal count data that cannot be negative.

For the empirical time series, linear trends in pH were relatively flat but slightly positive (Fig. S5A) whereas SST was steeper and had a positive slope (Fig. S6A). Trends were less pronounced in AQR than the other sites (Table S1). $ToE_{pH}$ ranged from 24-244 yrs (median=99 yrs) and $ToE_{SST}$ ranged from 6 to 3368 yrs (median=19 yrs) (Table S1, Table S2). ToE tended to be faster for SST than pH (excluding AQR which had an SST slope nearly equal to zero). Detrending the time series by removing the mean monthly effects did not improve the ToE estimates, because the detrending added an additional source of variability ($\bar{V}_I$) without greatly reducing the RMSE of the linear regressions (Figs. S5-S7).

ToE for the simulated data sets varied substantially across the scenarios evaluated and the time-series lengths used (Table S2). The range of $ToE_{SST}$ values was greater than for $ToE_{pH}$. Median $ToE_{SST}$ for the longer, 20-yr data sets was 84 yrs which was more than double the median $ToE_{SST}$ for the shorter, 6-yr datasets, but median $ToE_{pH}$ estimates were comparable among the two different time-series lengths.

## Literature Cited

Carter BR, Williams NL, Evans W, Fassbender AJ, Barbero L, Hauri C, Feely RA, Sutton AJ (2019) Time of Detection as a Metric for Prioritizing Between Climate Observation Quality, Frequency, and Duration. Geophysical Research Letters 46:3853–3861. https://doi.org/10.1029/2018GL080773

Keller KM, Joos F, Raible CC (2014) Time of emergence of trends in ocean biogeochemistry. Biogeosciences 11:3647–3659. https://doi.org/10.5194/bg-11-3647-2014

Table S1. Estimated time of emergence (ToE) for time-series of pH, sea surface temperature (SST), and sea surface salinity (SSS) for three sites (AQR, MKO, SBH). ToE was calculated using data that was not detrended (--) or detrended by month (mo.).

| | | ToE (yrs) | | |
|---|---|---|---|---|
| Variable | Detrending | AQR | MKO | SBH |
| ph | -- | 44.6 | 155.6 | 24.1 |
| ph | mo. | 64.8 | 243.6 | 133.4 |
| SST | -- | 3070.3 | 12.8 | 6.0 |
| SST | mo. | 3367.6 | 26.0 | 10.7 |
| SSS | -- | 184.2 | 11.4 | 2.6 |
| SSS | mo. | 58.6 | 18.2 | 4.3 |

Table S2. Summary of the estimated time of emergence (ToE) for empirical and simulated time-series of pH, sea surface temperature (SST). For simulated (sim.) data, values are for baseline parameter scenarios and intermediate long-term (50-yr) trends, with minimum and maximum values across all other scenarios in parentheses. For observed (obs.) empirical data, values are the median (min, max) across three sites.

| **Metric** | **Data** | **$ToE_{pH}$ (yrs)** | **$ToE_{SST}$ (yrs)** |
|---|---|---|---|
| Time of emergence for env. time series | obs. | 99 (24, 244) | 19 (6, 3368) |
| | sim. (6 yrs) | 63 (45, 84) | 40 (22, 102) |
| | sim. (20 yrs) | 61 (37, 88) | 84 (59, 116) |

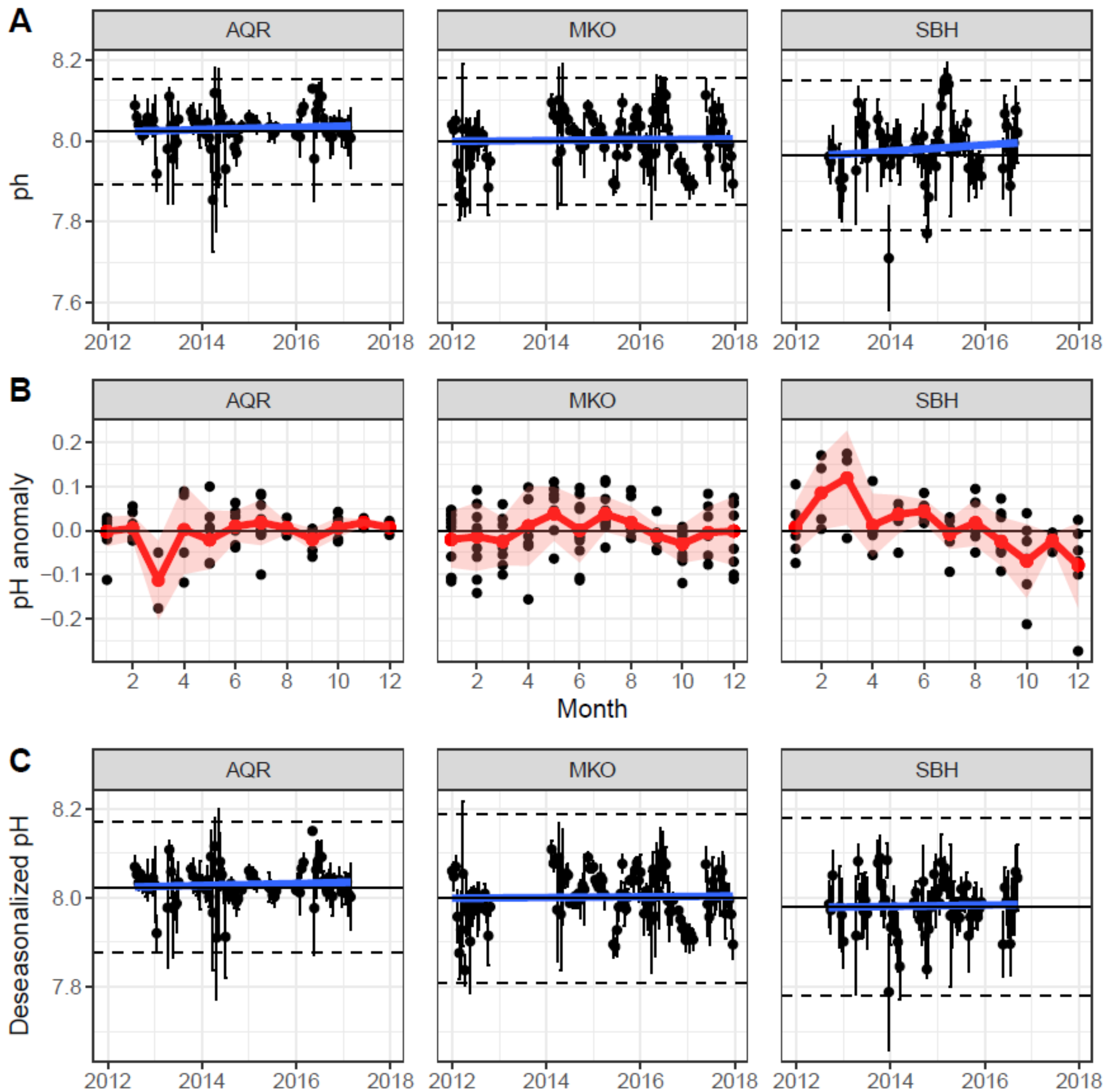


Fig. S5. Biweekly pH data from three different sites. (A) pH (+/- SD), (B) pH anomaly by month (red line = mean +/- SD), and (C) pH (+/- SD) after having the mean monthly anomaly removed. Sites are defined in Fig. 1 of the main paper. Blue lines are the linear regression fit to the plotted data. Horizontal lines in A and C represent twice the noise (+/-2N, dashed lines) around the first value of the linear regression (solid line). Time of emergence represents the number of years for the blue regression to exceed the background uncertainty (dashed line).

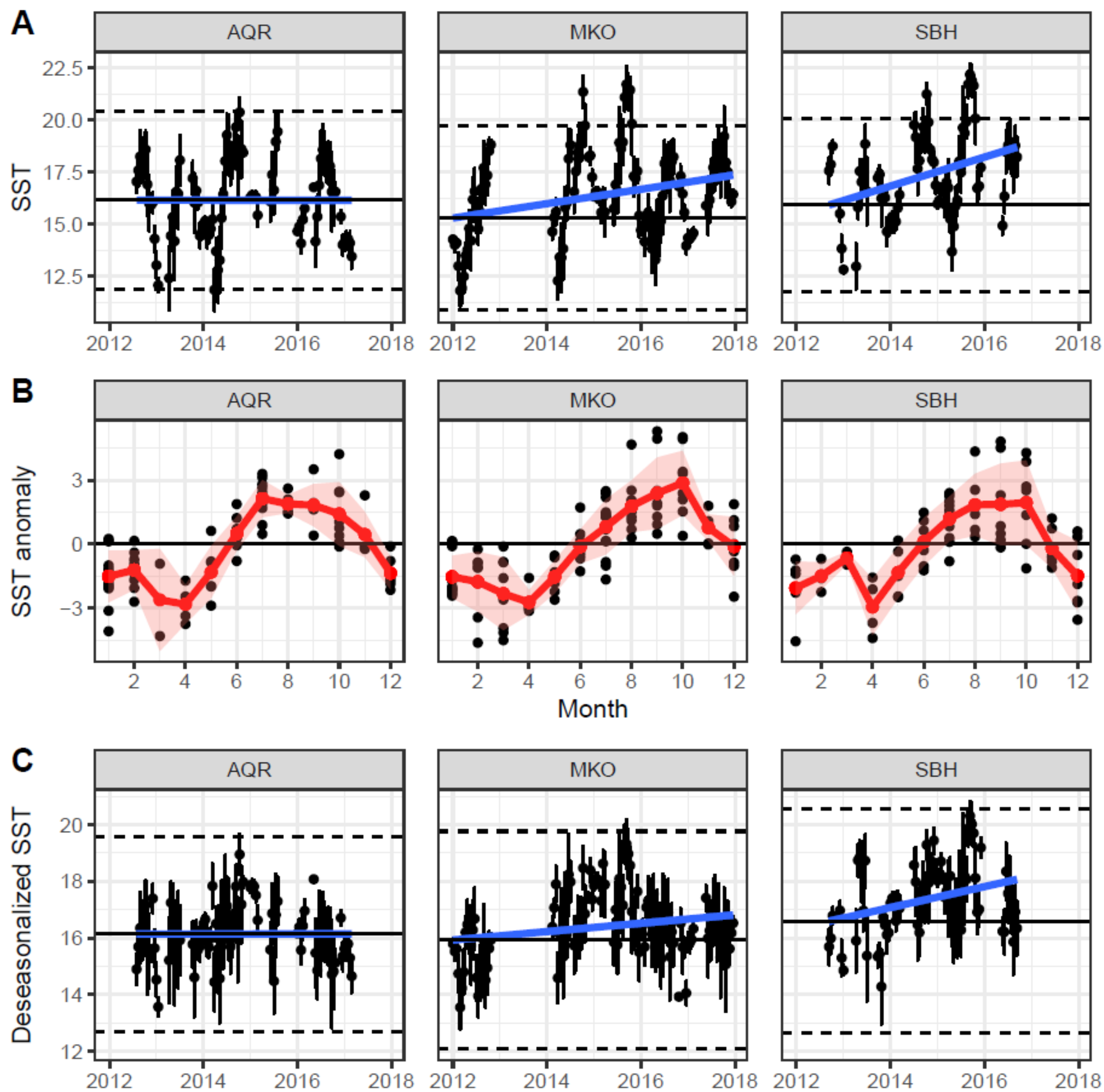


Fig. S6. Biweekly sea surface temperature (SST, °C) data from three different sites. (A) SST (+/- SD), (B) SST anomaly by month (red line = mean +/- SD), and (C) SST (+/- SD) after having the mean monthly anomaly removed. Sites are defined in Fig. 1 of the main paper. Blue lines are the linear regression fit to the plotted data. Horizontal lines in A and C represent twice the noise (+/- 2N, dashed lines) around the first value of the linear regression (solid line). Time of emergence represents the number of years for the blue regression to exceed the background uncertainty (dashed line).

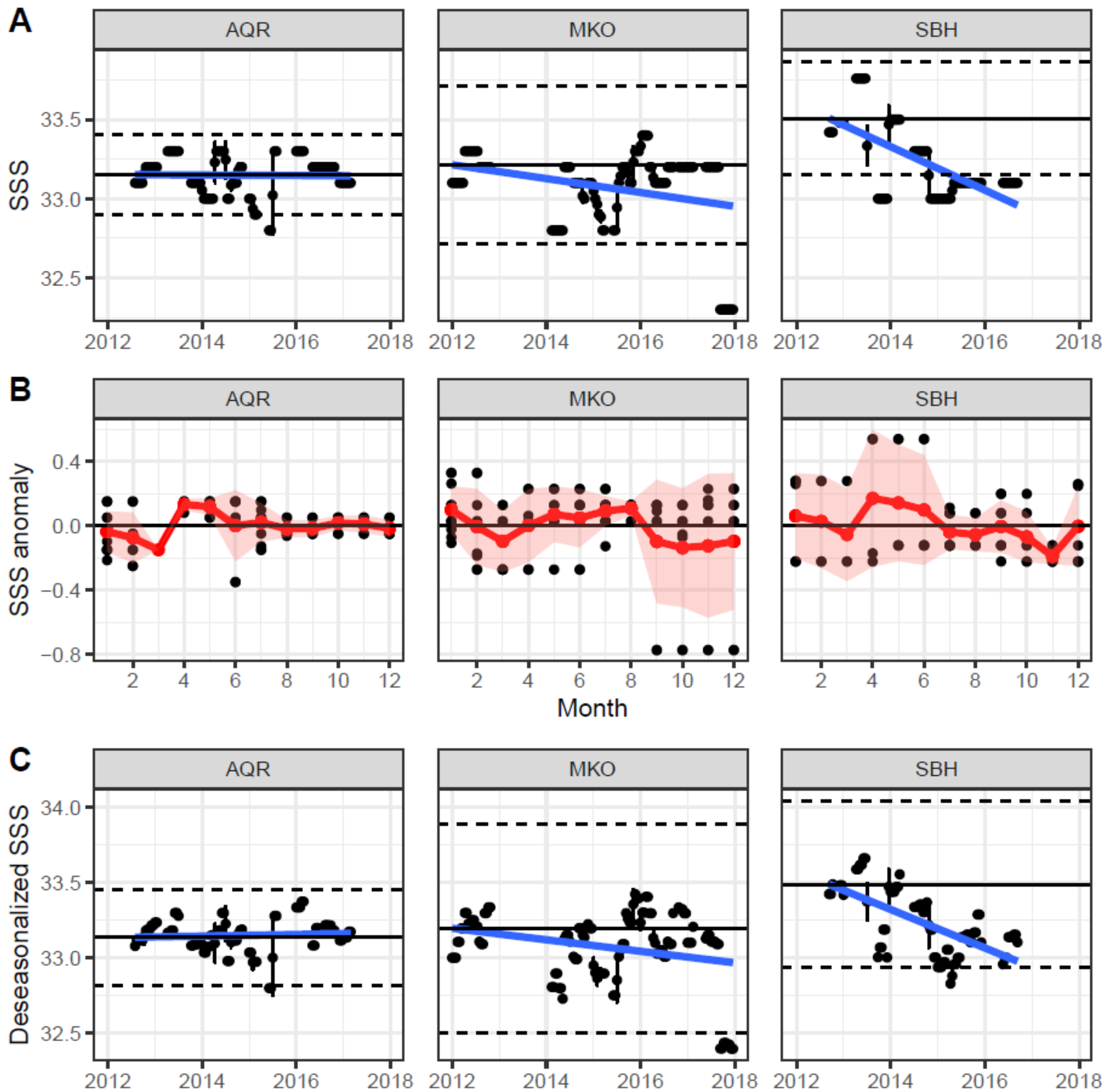


Fig. S7. Biweekly sea surface salinity (SSS, psu) data from three different sites. (A) SSS (+/- SD), (B) SSS anomaly by month (red line = mean +/- SD), and (C) SSS (+/- SD) after having the mean monthly anomaly removed. Sites are defined in Fig. 1 of the main paper. Blue lines are the linear regression fit to the plotted data. Horizontal lines in A and C represent twice the noise (+/-2N, dashed lines) around the first value of the linear regression (solid line). Time of emergence represents the number of years for the blue regression to exceed the background uncertainty (dashed line).