Additional Information on the Datasets and Modelling

All files and R scripts mentioned here are available to download from the GitHub depositary.

1. Indian Ocean Dipole – DMI Index

The DMI index is an indicator of the east-west temperature gradient across the tropical Indian Ocean, linked to the Indian Ocean Dipole or Zonal Mode. It is calculated as the difference of the WTIO and SETIO indices (Fig 1.1). This index was defined by Saji et al. (1999).

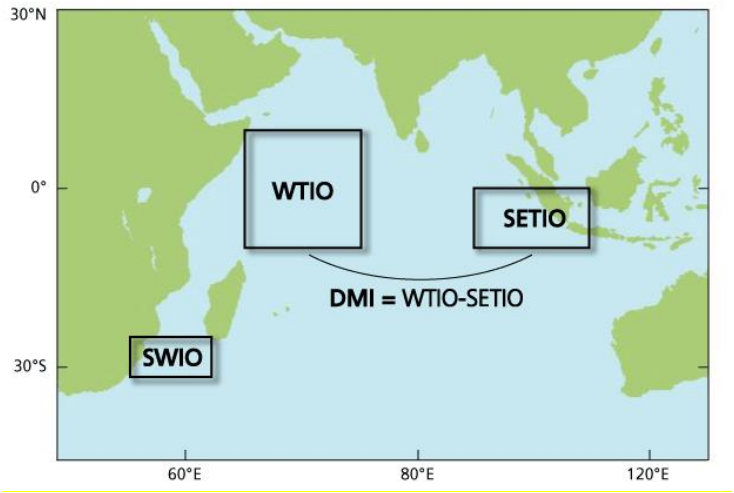

Fig S1.1. Components of the Indian Ocean Dipole Mode Index (DMI).

DMI data on the State of the Ocean Climate website[1] which were used in these analyses are based upon the Reynolds OIv2 SST analysis[2], made available by NOAA/ESRL. Data are weekly values from 1982 to the present day relative to a monthly climatological seasonal cycle based on the years 1982-2005. The monthly climatology is then linearly interpolated to produce a weekly anomaly series. The uncertainty of the index is estimated to be 0.1921°C over the period 1982-present.

Plotting and data manipulation were carried out in R (R Core Team 2019) using the script file "*Analysis of DMI from NOAA NetCDF data – Figure 1B.R*" and the NetCDF file "*dmi to 13 Jan 2021.nc*". Both are included in the GitHub repository.

[1] https://stateoftheocean.osmc.noaa.gov/sur/ind/dmi.php

[2] The optimum interpolation (OI) sea surface temperature (SST) analysis is produced weekly on a one-degree grid. The analysis uses in situ and satellite SSTs plus SSTs simulated by sea ice cover. Before the analysis is computed, the satellite data is adjusted for biases using the method of Reynolds (1988) and Reynolds and Marsico (1993). A description of the OI analysis can be found in Reynolds and Smith (1994). The bias correction improves the large-scale accuracy of the OI. In November 2001, the OI fields were recomputed for late 1981 onward. The new version is referred to as OI.v2.
The most significant change for the OI.v2 is the improved simulation of SST observations from sea ice data following a technique developed at the UK Met Office. This change has reduced biases in the OI SST at higher latitudes. Also, the update and extension of COADS has provided us with improved ship data coverage through 1997, reducing the residual satellite biases in otherwise data sparse regions. For more details, see Reynolds, et al (2002).

2. El Niño-Southern Oscillation (ENSO)

https://stateoftheocean.osmc.noaa.gov/sur/pac/nino34.php

We used the Nino3.4 index as the most representative index of ENSO (Bamston et al. 1997). The Niño3.4 Sea Surface Temperature (SST) anomaly index is an indicator of central tropical Pacific El Niño conditions. It is calculated with SSTs in the box 170°W-120°W, 5°S-5°N (Fig 2.1).

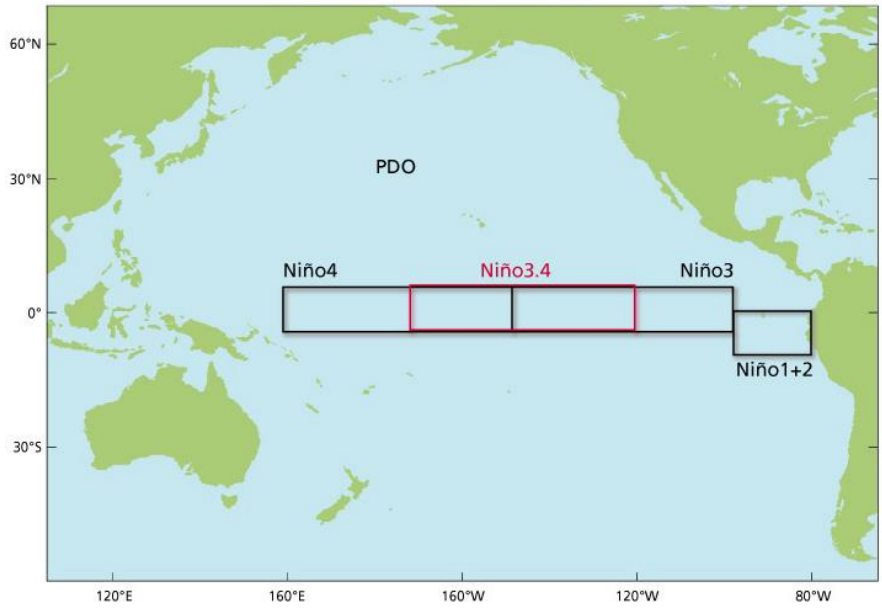

Fig S2.1. Location of boxes for Nino 3.4 Index.

Nino 3.4 data on the State of the Ocean Climate website[3] which were used in these analyses are based upon the Reynolds OIv2 SST analysis[4], made available by NOAA/ESRL. The anomaly is calculated relative to a monthly climatological seasonal cycle based on the years 1982-2005. The monthly climatology is then linearly interpolated to determine weekly anomalies. Spatial averaging of the gridded analysis is weighted by surface area.

Plotting and data manipulation were carried out in R (R Core Team 2019) using the script file "*Analysis of Nino 3.4 from NOAA NetCDF data – Figure 1B.R*" and the NetCDF file "Nino 3.4 *to 13 Jan 2021.nc*".

Monthly means were computed for both Nino 3.4 and DMI to align to data for absolute sea level for use in modelling.

[3] https://stateoftheocean.osmc.noaa.gov/sur/pac/nino34.php

[4] The optimum interpolation (OI) sea surface temperature (SST) analysis is produced weekly on a one-degree grid. The analysis uses in situ and satellite SSTs plus SSTs simulated by sea ice cover. Before the analysis is computed, the satellite data is adjusted for biases using the method of Reynolds (1988) and Reynolds and Marsico (1993). A description of the OI analysis can be found in Reynolds and Smith (1994). The bias correction improves the large-scale accuracy of the OI. In November 2001, the OI fields were recomputed for late 1981 onward. The new version is referred to as OI.v2.
The most significant change for the OI.v2 is the improved simulation of SST observations from sea ice data following a technique developed at the UK Met Office. This change has reduced biases in the OI SST at higher latitudes. Also, the update and extension of COADS has provided us with improved ship data coverage through 1997, reducing the residual satellite biases in otherwise data sparse regions. For more details, see Reynolds, et al (2002).

3. Satellite Altimetry of absolute sea level

CMEMS 008-057[5] satellite altimetry data [Sea Level Anomaly (SLA)] was used for examining regional patterns of temporal changes in absolute sea-level and for analysing the time series at a specific location. This dataset is DUACS[6] delayed-time altimeter gridded data of sea surface heights over the global ocean, using a stable two-satellite constellation. The product is intended, inter alia, for monitoring long-term sea-level change. It is produced and distributed by the Copernicus Climate Change Service[7] (C3S) as downloadable NetCDF4 files. The present time span of the data is 1 Jan 1993 to 7 March 2020. Data are daily values[8].

Data were downloaded for the latitude and longitude grid square 96°-100°E, 6°-10°N which includes the study site at Cape Panwa, Phuket in NetCDF format – file "*dataset-duacs-rep-global-merged-twosat-phy-l4_1608108067774.nc*".

Grid Square of sea-level rise

To produce the area grid of sea-level rise, an Ordinary Least Squares regression (OLS) was run for each 0.25 x 0.25 degree Latitude/Longitude grid square (n=289) using the R core function 'lm'. Because the purpose of the gridded data was simply to illustrate how sea-level rise varies across the area, no further checking of the assumptions, nor of the goodness of fit were required. A specific grid point was then chosen for further analysis. The R script file "*Phuket CMEMS 008-057 NetCDF data – Grid – Figure S2.R*" contains the code for this.

4. Modelling patterns in the sea level time series for the grid square, 7.125°N 98.125°E.

From past experience, patterns in the sea level time series in the Andaman Sea area are known to be influenced by several factors, namely (1) long-term rise due to warming of the oceans, (2) seasonality, represented by monthly variation in mean sea level, and (3) the Indian Ocean Dipole (IOD). It was also hypothesised that ENSO might exert an influence, and that any change in the timing of the reversing monsoon in this area might also be reflected in long-term changes to the monthly variation pattern (i.e., an interaction between time and seasonality).

It was also assumed that variables (1) might not act linearly, (2) would likely be additive, and that given that this is a time series it was probable that temporal autocorrelation would be present. Accordingly, generalized additive mixed models (GAMM) were chosen to incorporate the non-linear form and to include an autoregressive term.

The R package 'mgcv' (Wood 2017) was used for GAMM model evaluation. Models were built from the logical predictors using (1) decimal years [a continuous variable e.g., 1982.4], (2) month number [1-12], (3) Dipole Mode Index [a continuous variable], and (4) the Nino 3.4 index [continuous variable]. Covariates were modelled as fixed effects (i.e. affecting all the data). Non-significant terms were

[5] Copernicus Marine Environment Monitoring Service: https://resources.marine.copernicus.eu/?option=com_csw&task=results?option=com_csw&view=details&product_id=SEALEVEL_GLO_PHY_CLIMATE_L4_REP_OBSERVATIONS_008_057

[6] Data Unification and Altimeter Combination System. The DUACS system is the CNES/CLS processing system that provides satellite altimeter sea level products. It is used for the operational production of sea level products for the Marine (CMEMS) and Climate (C3S) services of the E.U. Copernicus program, for the processing of the Sentinel-3 products on behalf of EumetSat and for the production of demonstration and pre-operational products on behalf of the CNES French space agency.

[7] https://climate.copernicus.eu/

[8] Note when specifying lat and long in CMEMS for longitude use eg 100.125 to include grid square for 100.

removed from models. The response variable (SLA) and the DMI covariate were both already in the form of an anomaly and therefore represented centred (on zero) values.

All analyses were conducted in R using the R script file "*Phuket CMEMS 008-057 NetCDF data – GAMM models.R"* which contains further details and notes on use. Values for monthly DMI and Nino 3.4 obtained from the earlier analyses of these indices had been stored in the CSV file "*DMI Nino 3.4 monthly to 14 Mar 2020.csv*" and were input to the R code where indicated. As indicated above, the data for satellite altimetry was input from the NetCDF file "*dataset-duacs-rep-global-merged-twosat-phy-l4_1608108067774.nc*".

Data Exploration

Data exploration was applied following the protocols in Zuur et al. (2010). The presence of possible outliers was investigated using Cleveland dotplots or boxplots.

SLA data plotted by month (Fig S4.1) reveals a seasonal pattern with lowest sea level in Feb/Mar each year. Negative outliers (sea-level depressions) occurred in 1994, 1997, 1998, 2003, and 2006. Positive outliers (sea-level elevation) were present in 2010, 2013, 2014, and 2016. Since these years are known to have been associated with large positive or negative IODs, the outliers were retained in the dataset as valid readings.

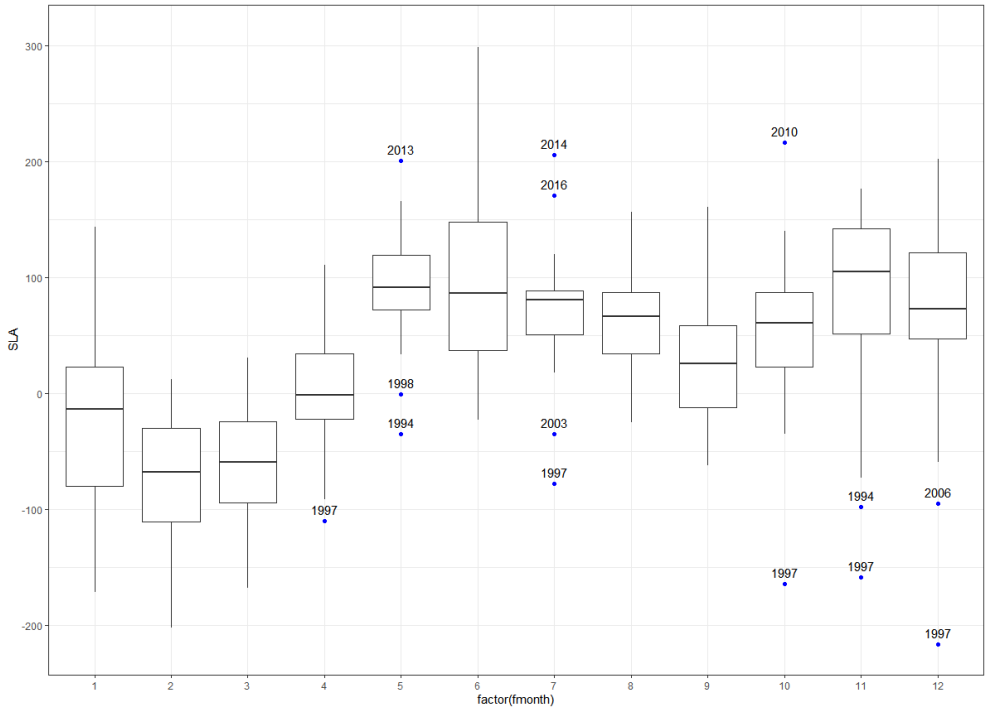

Fig S4.1. Boxplot of monthly SLA data (1993-2020) for the grid square, 7.125°N 98.125°E, ~89km from Cape Panwa, Phuket. Outliers, in blue, are identified by the year in which they occur. Outliers are 1.5 times the inter quartile range above (3rd Quartile) or below (1st Quartile).

The observed pattern of seasonality confirmed that a term (month) for this should be included in the model.

Collinearity between the potential covariates was assessed using multi-panel scatterplots, Pearson correlation coefficients and variance inflation factors (VIF). VIFs were computed using the 'corvif' function (Zuur et al. 2009).

There was no collinearity apparent between the chosen covariates (DMI, month, time, and Nino 3.4) (Fig S4.2). This was confirmed by variance inflation factors (VIF) which were approximately 1.0 in each case. A cut off value of 3 for the VIF was used.

Covariate	GVIF
DMI	1.190
month	1.048
time	1.01
Nino3.4	1.163

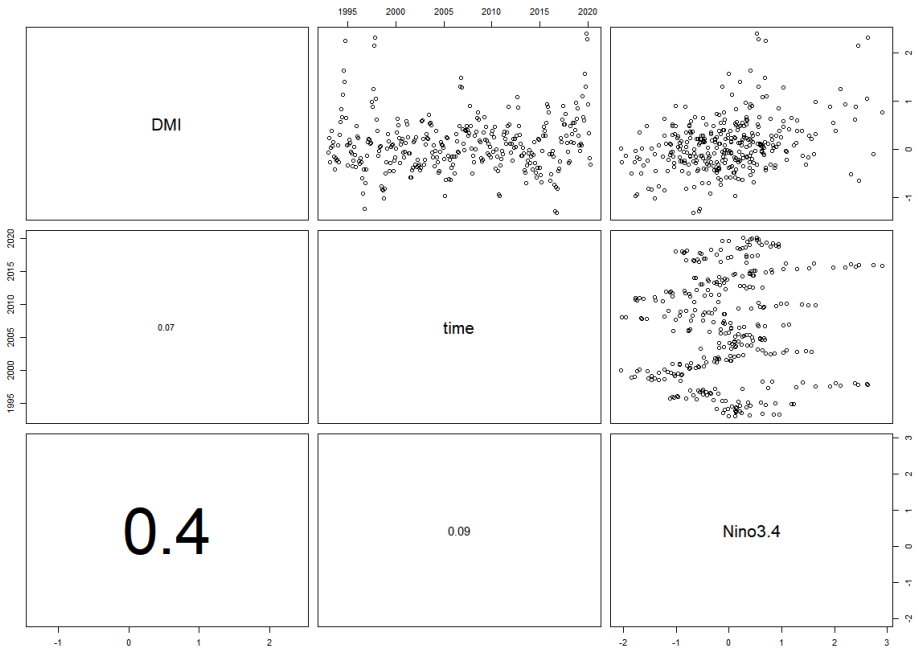

Fig S4.2. Pair plots for covariates (top right) with Pearson correlation coefficients (bottom left). Monthly SLA data (1993-2020) for the grid square, 7.125°N 98.125°E, ~89km from Cape Panwa, Phuket.

Datasets were checked for missing data. The relationships between SLA and individual covariates were visualised using multi-panel scatterplots (Fig S4.3).

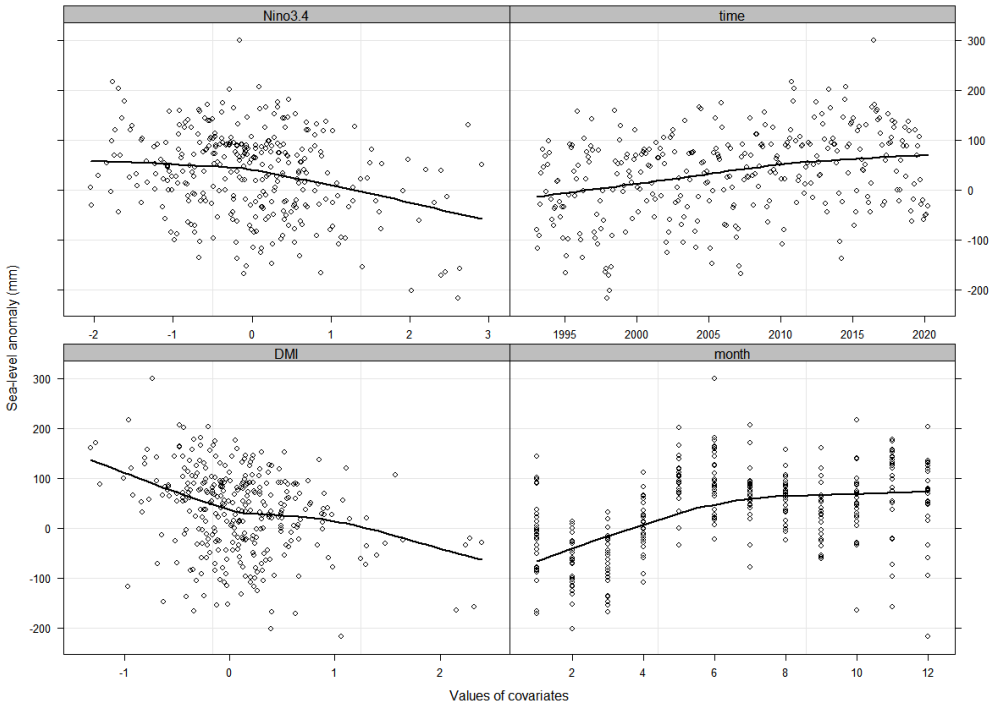

Fig S4.3. Covariates plotted against SLA. Monthly SLA data (1993-2020) for the grid square, 7.125°N 98.125°E, ~89km from Cape Panwa, Phuket. A Loess smoother has been added to each plot to aid interpretation.

Model choice and selection

Exploratory ACF/PACF plots identified autocorrelation in the sea level data and were used to indicate the autoregressive AR() term which would probably be needed to remove this. A first order AR(1) term was found to be suitable. For this term, month was nested within year (i.e. the covariate determining the ordering of the residuals was 'month' and the correlation applied within measurements made in one year).

The smoothing method REML (restricted maximum likelihood) was chosen since this is known to give stable, reliable results. 'mgcv' was initially allowed to run in its automated mode to select appropriate smoothing parameters and basis functions. If necessary specific values for these parameters were later added to the model.

Thin plate spline (tp) basis functions were used for time and DMI variables. The seasonal term (month) was initially modelled using cubic cycle (cc) for the repeating seasonality but where patterns still remained in the residuals, month was modelled as a factor to allow different intercepts for each month.

An interaction between time and seasonality (i.e., the monthly seasonality changing over the timescale of the times series) was investigated by the inclusion of an interaction term using a tensor product interaction to account for the different scales of the time and seasonal components. The interaction term was however non-significant and was dropped from further models.

Several GAMM models, all of which included terms for Dipole Mode Index (DMI), Nino 3.4, time, and seasonality (month) were computed. In the 'M2aa' model DMI and Nino 3.4 were included as smoothing terms but both had effective degrees of freedom (edf) close to 1, indicating that these could be linear terms. The model 'M2aa_linear_AR1' listed below is the same model as M2aa but with these terms linear.

Model selection was based upon AICc (AIC with small sample bias adjustment). Adjusted R^2 values were used to assess the amount of variation in the sea-level time series accounted for by the model and individually for each additive term.

A range of models were compared and the best fit chosen based upon lowest AICc using the package 'MuMIn' and the 'model.sel' function (Barton 2020).

	X	family	df	logLik	AICc	delta	weight
M2aa_linear_AR1$lme	+	gaussian(identity)	18	-1642.71	3323.64	0.000	0.969
M2aa$lme	+	gaussian(identity)	20	-1644.30	3331.34	7.691	0.021
M2x$lme	+	gaussian(identity)	31	-1631.97	3332.66	9.013	0.011
M2$lme	+	gaussian(identity)	10	-1695.21	3411.12	87.479	0.000

The optimum model (M2aa_linear_AR1) appears first in the list and explained 73.1% of the variation in SLA. The model included fixed effects of DMI and Nino3.4 (both linear terms), and time (smoothing term), with a seasonality (month) term as a factor (separate intercepts for each month).

R code for the model fit was:

M2aa_linear_AR1 <- gamm(SLA~ s(time) + DMI + Nino3.4 + fmonth, data = monthly_SLA_DMI, correlation = corAR1(form = ~ month | year), method="REML", select=FALSE)

All model terms were significant. The effective degrees of freedom for Time (edf=4.477) indicated curvature in this term.

```
Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  35.6151     3.3129  10.750  < 2e-16 ***
DMI         -33.0013     5.8648  -5.627 4.12e-08 ***
Nino3.4     -26.5899     3.7144  -7.158 5.99e-12 ***
fmonth.L    134.1776    10.6142  12.641  < 2e-16 ***
fmonth.Q    -87.4493     9.9901  -8.754  < 2e-16 ***
fmonth.C      0.8516     9.3104   0.091 0.927177
fmonth^4     81.6286     8.7996   9.276  < 2e-16 ***
fmonth^5    -85.7123     8.2428 -10.398  < 2e-16 ***
fmonth^6    -23.1191     7.7878  -2.969 0.003226 **
fmonth^7     19.8846     7.3602   2.702 0.007281 **
fmonth^8    -16.8876     7.0400  -2.399 0.017043 *
fmonth^9     -0.7983     6.7524  -0.118 0.905971
fmonth^10    23.4843     6.5191   3.602 0.000367 ***
fmonth^11     2.8954     6.2939   0.460 0.645818
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
          edf Ref.df     F p-value
s(time) 4.477  4.477 22.14  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.731
  Scale est. = 1947.6    n = 327
```

Phi=0.305777

fmonth p<2e-16

<u>Model checking</u>

Models were checked using the 'mgcv' function 'gam.check' for convergence, and the appropriate number of basis functions. The 'concurvity' function was used to check for correlation between variables. Model validation also included the following steps:

1. To verify homogeneity of variance, residuals were plotted versus the fitted values.

2. To verify model misfit (or independence), residuals were plotted versus each covariate in the model.
3. To verify independence (repeated measurements over time), auto-correlation plots were created from the residuals.
4. To verify the normality assumption, a histogram of the residuals was created.
5. The model was inspected for influential observations.

```
> gam.check(M2aa_linear_AR1$gam)

'gamm' based fit - care required with interpretation.
Checks based on working residuals may be misleading.
Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

          k'  edf k-index p-value
s(time) 9.00 4.48    0.74  <2e-16 ***
---
```

Residuals were normal, homoscedastic and the AR(1) term removed the autocorrelation (Fig S4.4). There was no problematic concurvity (worst case 0.21).

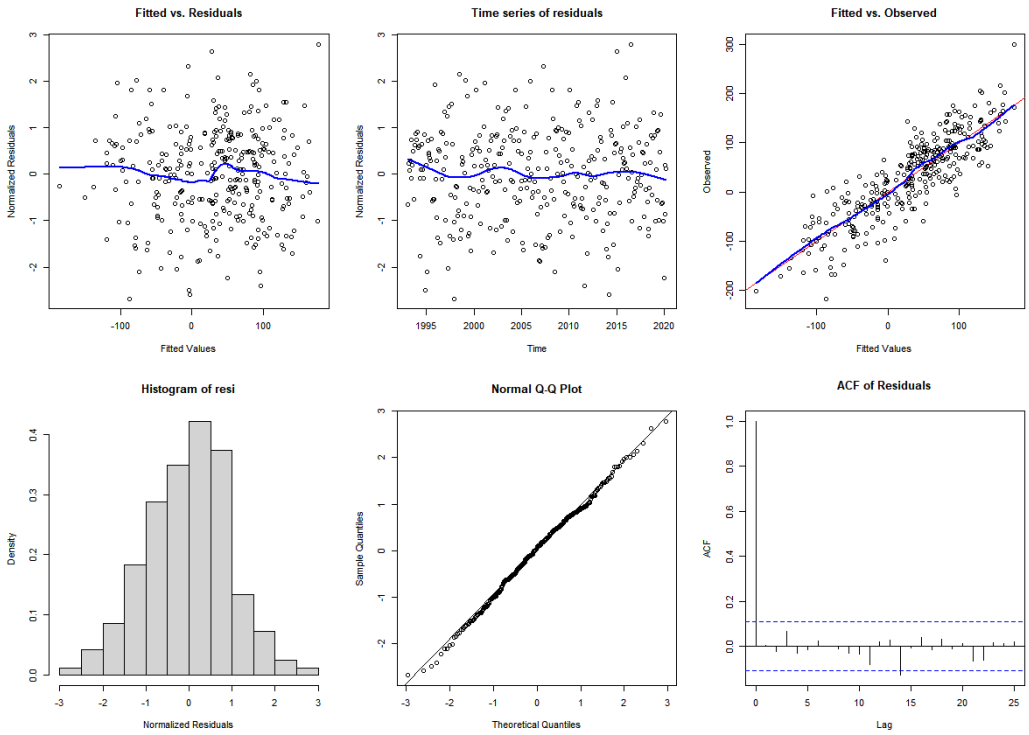

Fig S4.4. Diagnostic plots for the final model.

Plots of the model residuals against each covariate (Fig S4.5) indicated that there were no patterns remaining in the residuals.

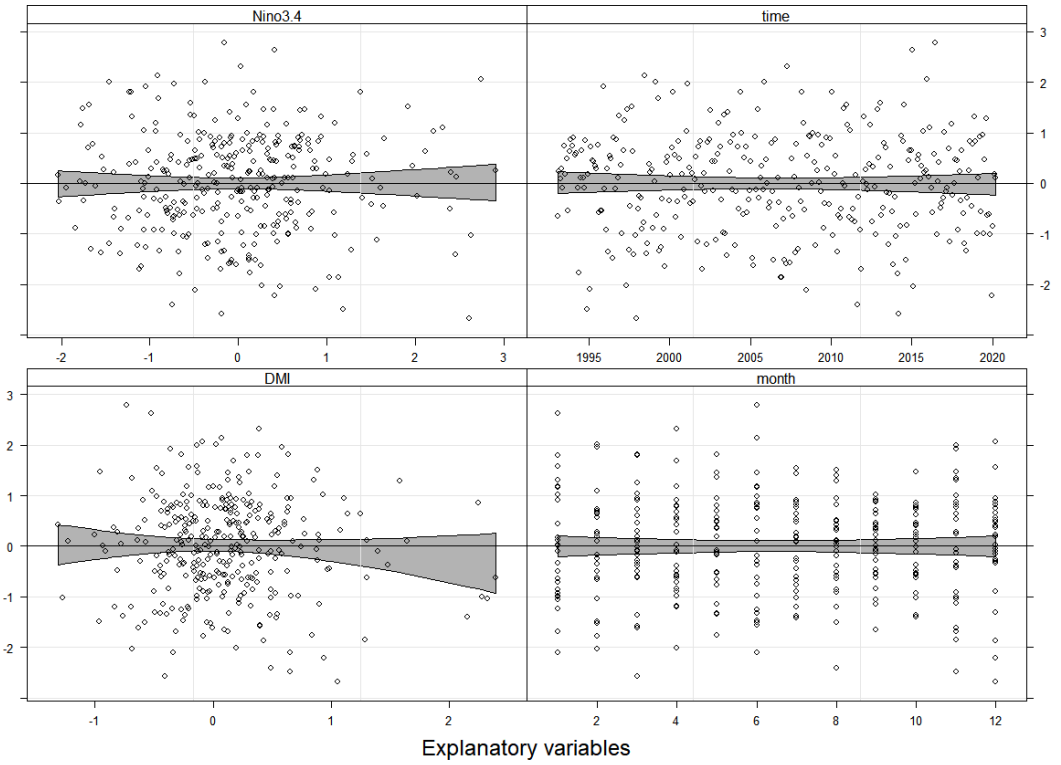

Fig S4.5. Plots of residuals versus covariates. Smoother and 95% point-wise confidence intervals (grey shade). The confidence intervals contain 0 for all covariate values indicating no significant patterns.

Finally, a conditional boxplot of the residuals by month (Fig S4.6) confirmed no significant residual month effect.

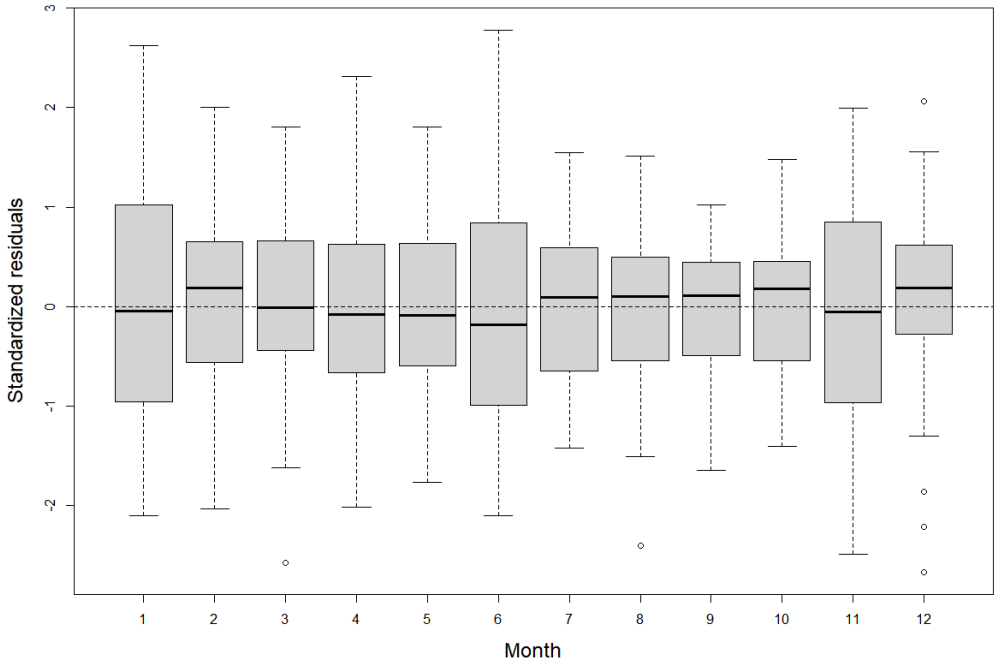

Fig S4.6. Conditional boxplots of residuals grouped by month.

Visualising the model output

As well as inspecting model output in the form of partial plots, the overall model and individual terms were extracted and plotted using the 'predict' function in 'mgcv'.

To represent the output from the model the effect of each term was examined separately as well as the overall model. Since the model is additive, the overall model is a summation of each term (time, DMI, month, Nino 3.4) together with the intercept.

Partial effect plots for each model term (Fig S4.7) show how each term separately contributes to the model whilst the other covariates are held at their median value. The relationship with Nino 3.4 and DMI were linear.

The linear relationship for DMI was -33 ± 5.9 mm DMI^{-1} $p<0.0001$, and is consistent with what is known about the effect of DMI on sea level at Phuket, where positive values of the DMI result in sea-level depression and vice versa (Brown et al. 2002; Brown et al. 2011). The trend with Nino 3.4 is also negative –26.6 ± 3.7 mm $Nino3.4^{-1}$ $p<0.0001$.

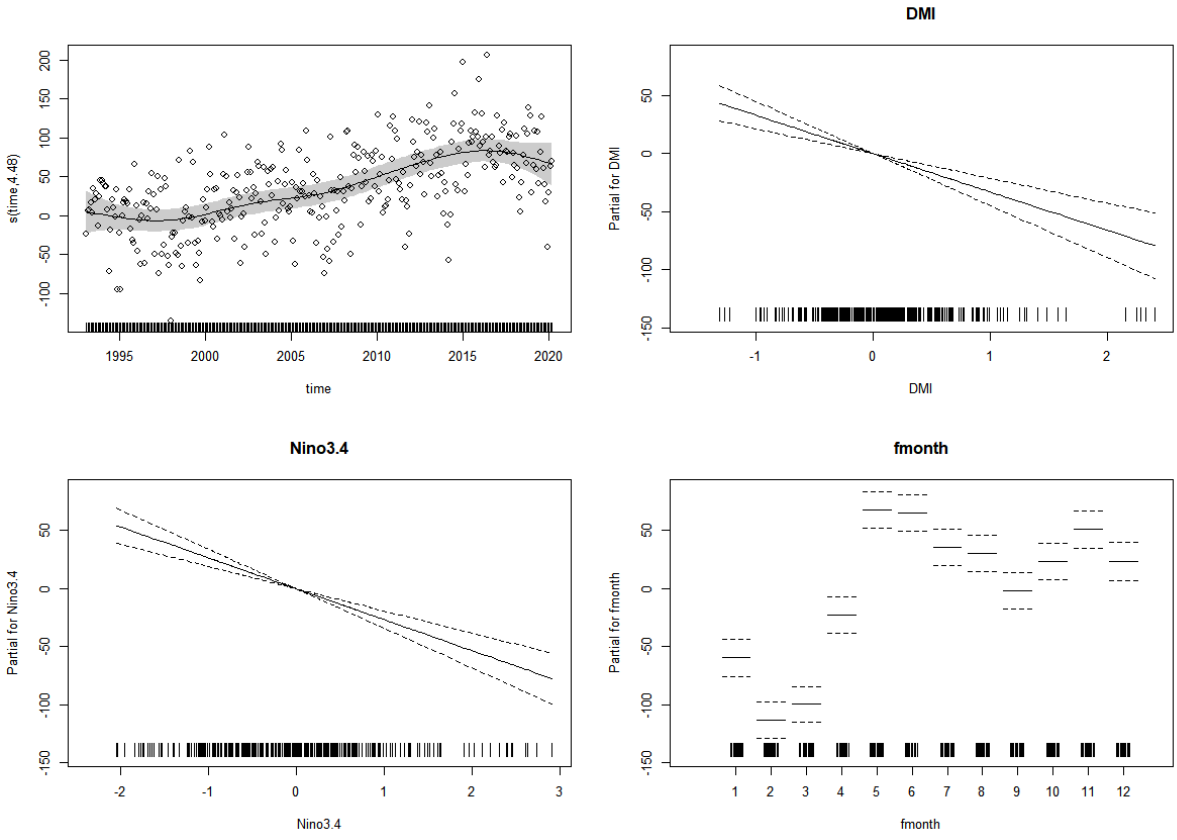

Fig S4.7. Partial effect plots for the GAMM model. (1) non-linear relationship between time and SLA (2) linear relationship between DMI and SLA (3) linear relationship between Nino 3.4 and SLA (4) Month as a factor. Residuals plotted as open circles. Partial effects include model intercept. Shaded areas/ dotted lines are 95% CI.

Individual Model Terms

The overall model fit for all terms is shown in Fig S4.8. Separate fits for each of the terms are shown in Figs S4.9-4.14. The model explained 73.1% of the variation in the data with the seasonality (fmonth) accounting for ~42.4%, time ~12.8%, DMI ~8.2%, and Nino3.4 ~6.4%.

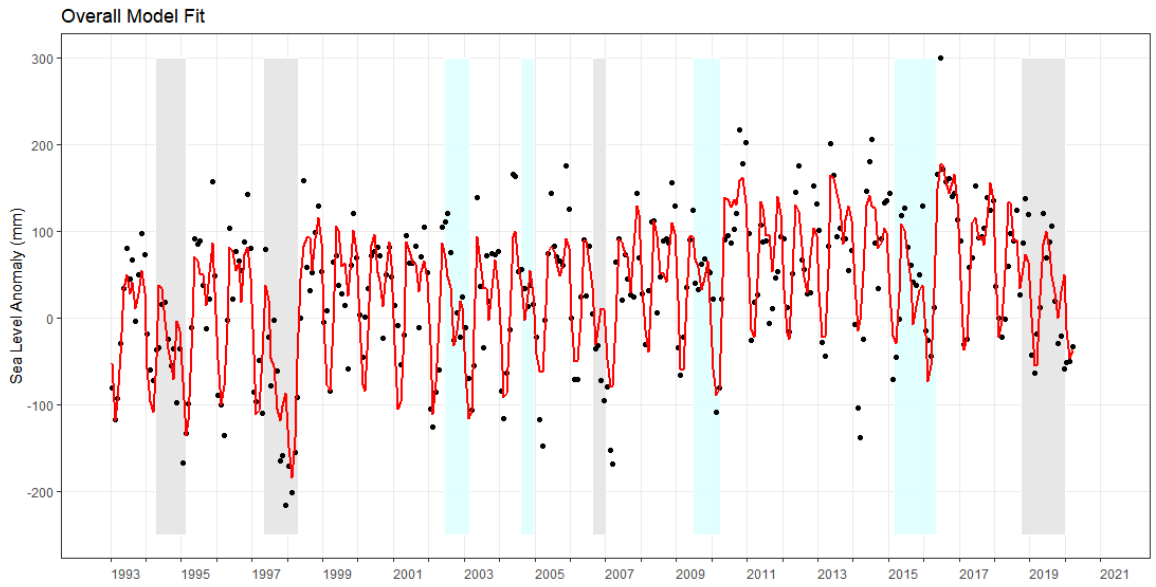

Fig S4.8. Monthly mean SLA data (black dots) with overall model fit (red line). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Nino (Nino3.4 > 0.5 °C); cyan shading El Nino only.

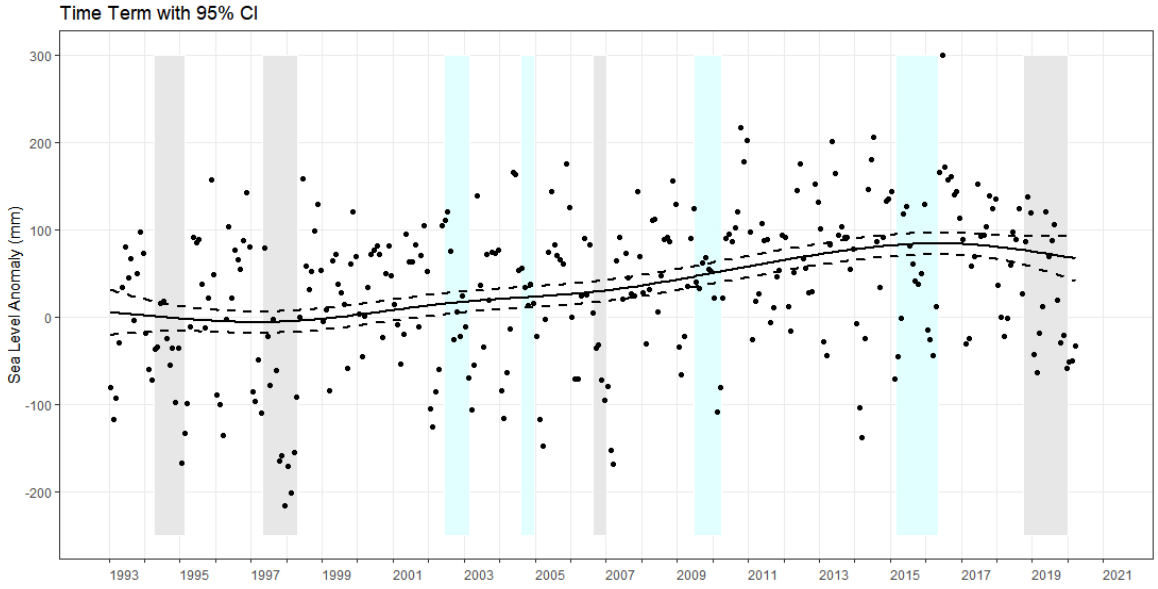

Fig S4.9. Monthly mean SLA data (black dots) with fit of the term for time (black line) and 95% confidence intervals (dotted lines). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Nino (Nino3.4 > 0.5 °C); cyan shading El Nino only.

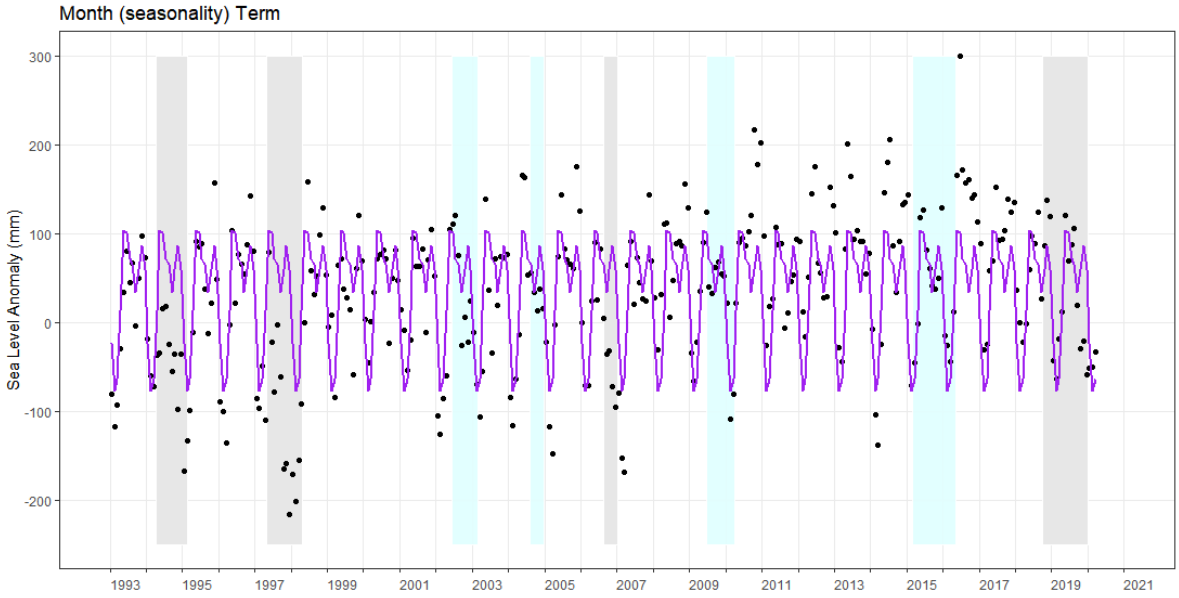

Fig S4.10. Monthly mean SLA data (black dots) with fit for seasonality term (month) (purple line). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Nino (Nino3.4 > 0.5 °C); cyan shading El Nino only.

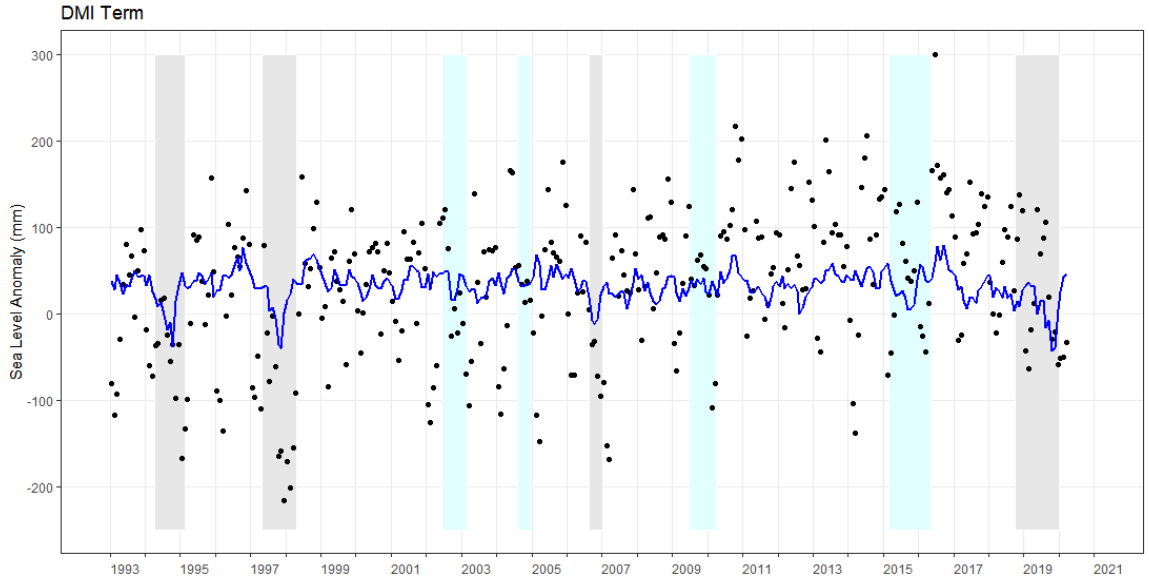

Fig S4.11. Monthly mean SLA data (black dots) with fit for DMI term (blue line). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Nino (Nino3.4 > 0.5 °C); cyan shading El Nino only.

The fit for DMI (Fig S4.11) shows that this term has identified the large *p*IOD in 1994, 1997, the smaller *p*IOD in 2006, and the large *p*IOD in 2019. In each case the plot shows the depression in SLA which accompanies the *p*IOD. The present dataset ends in March 2020.

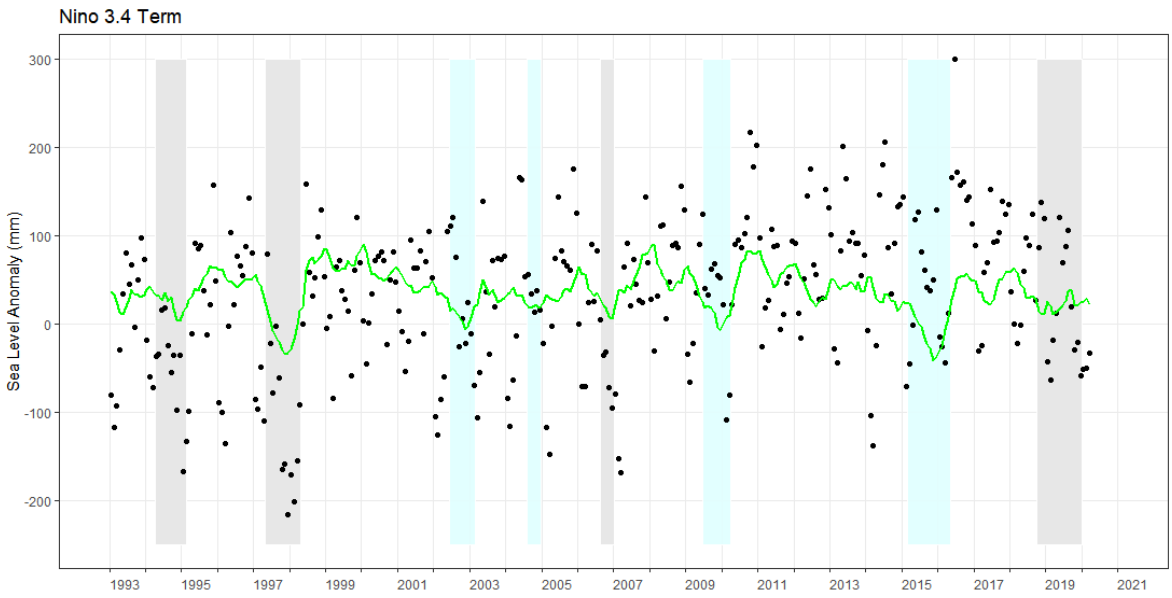

Fig S4.12. Monthly mean SLA data (black dots) with fit for Nino 3.4 term (green line). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Nino (Nino3.4 > 0.5 °C); cyan shading El Nino only.

The fit for Nino 3.4 (Fig S4.12) identifies the contribution to the sea-level depressions in 1994, 1997, 2002, 2006, 2009, 2015 and a relatively small contribution in 2019.

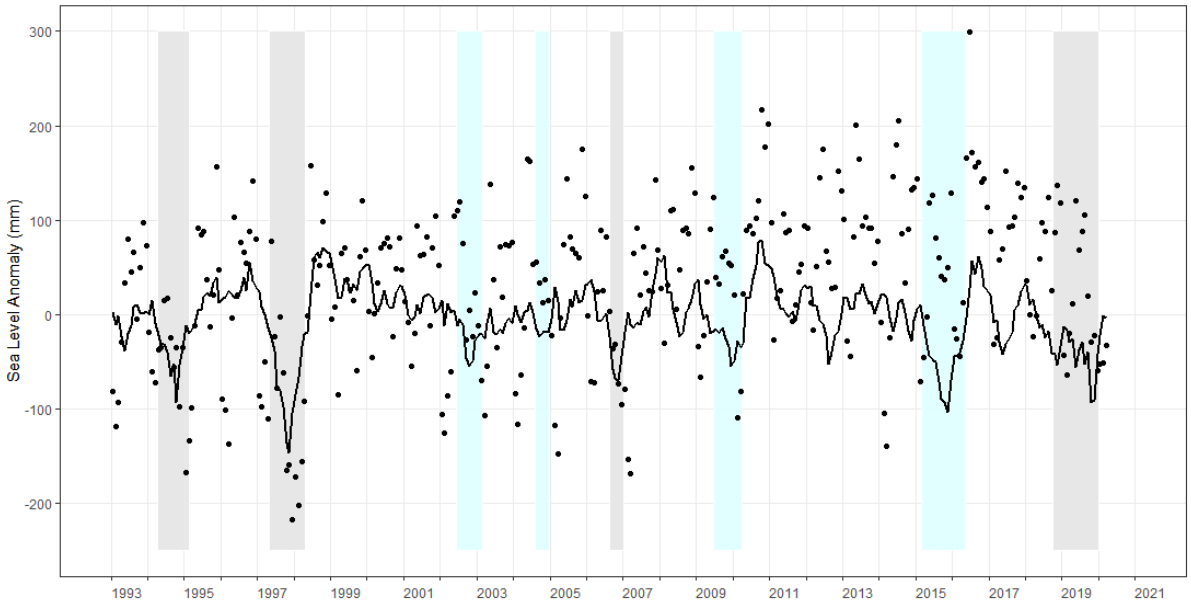

Fig S4.13. Monthly mean SLA data (black dots) with DMI and Nino 3.4 predicted values combined. Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Nino (Nino3.4 > 0.5 °C); cyan shading El Nino only.

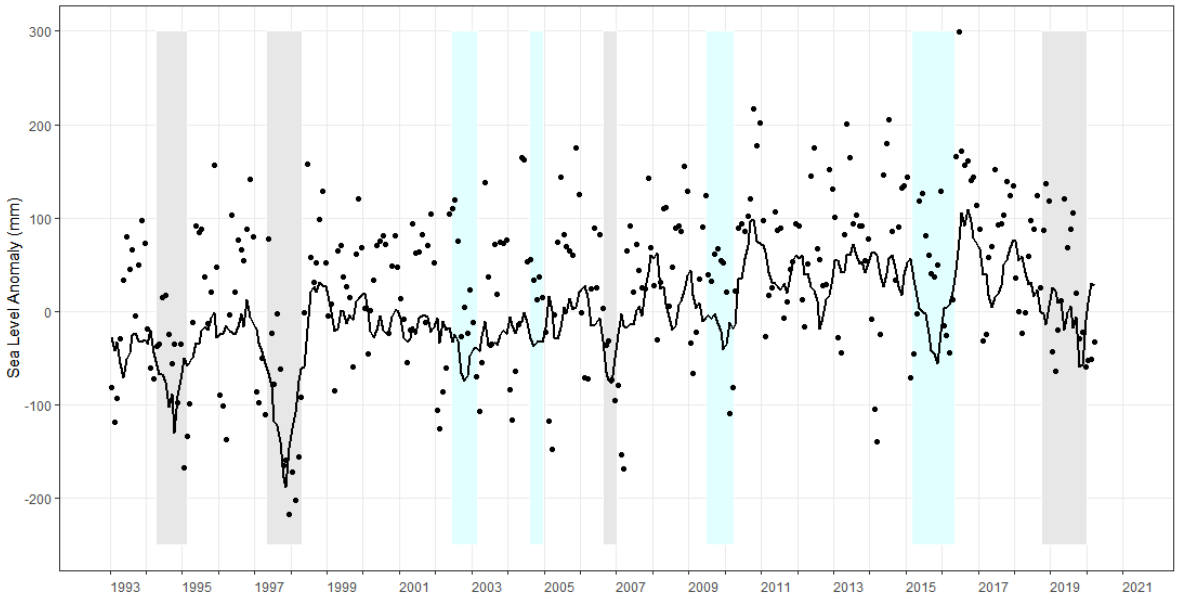

Fig S4.14. Monthly mean SLA data (black dots) with DMI, Nino 3.4 and time terms combined (i.e. the overall model with seasonality removed. Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 ^{o}C) and El Nino (Nino3.4 > 0.5 ^{o}C); cyan shading El Nino only.

The plot of the overall model with seasonality removed (Fig S4.14) shows that the increase in sea level over time reduces the magnitude of negative sea-level anomalies due to IOD and El Nino and vice versa for positive anomalies (compare with Fig S4.13).

5. Modelling patterns in the Hotspot series for the grid square adjacent to Cape Panwa.

Hotspot data from the Coral Reef Watch 5km satellite coral bleaching heat stress monitoring products (Version 3.1) (Liu et al. 2014) were downloaded for a 5km grid square 7.825°N 98.475°E, 7.8 km to the east of the study site at Cape Panwa, Phuket covering the time period 1 July 1987 to 13 January 2021. Data are daily values of Hotspot, which is the amount by which the daily mean satellite sea surface temperature is above the highest climatological monthly mean in °C.

The daily data for Hotspot for the period 1 May 1985 to 31 Dec 2020 is stored in the Excel File "*CRW data for Phuket 5km square – 1 May 1985 to 13 Jan 2021.csv*".

R code for the analysis is in file "*Phuket CRW data – GAMM models.R*".

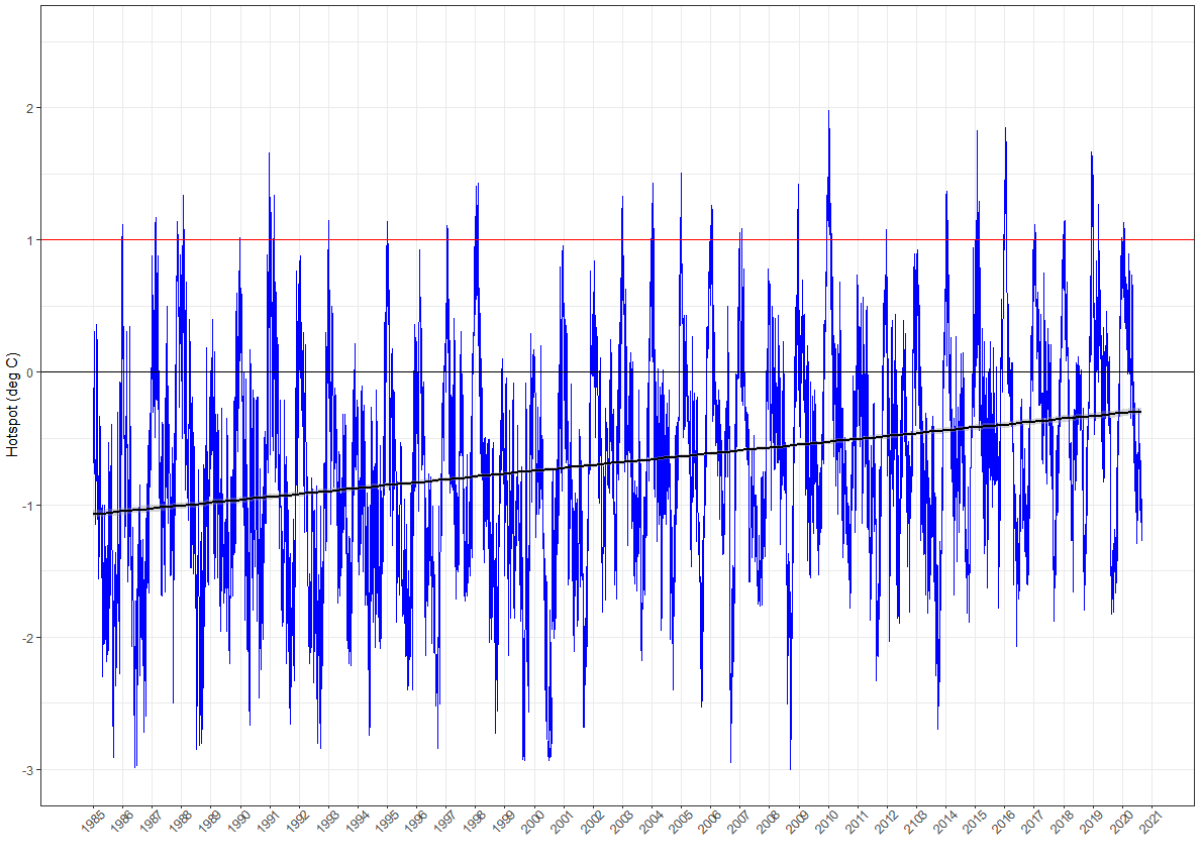

Fig S5.1. Plot of Hotspot daily data 1 May 1985 to 31 Dec 2020 for the grid square, 7.825°N 98.475°E, ~7.8km to the east of Cape Panwa, Phuket. Red line is +1°C above which Hotspots generate Degree Heating Weeks (DHW). Regression line is Ordinary Least Squares fit 0.0222°C yr^{-1}, 95% CI 0.0209-0.0236.

The potential relationships of IOD and ENSO with Hotspot data were investigated. The hypothesis being that either or both covariates influence the appearance of Hotspots >1°C. Terms selected were: (1) time (i.e. long-term temperature rise due to warming of the oceans), (2) seasonality, represented by monthly variation in Hotspot data, (3) the Indian Ocean Dipole (IOD), and (4) ENSO. Any change in the timing of the reversing monsoon in this area might also be reflected in long-term changes to the monthly variation pattern (i.e., an interaction between time and seasonality).

It was also assumed that variables (1) might not act linearly, (2) would likely be additive, and that given that this is a time series it was probable that temporal autocorrelation would be present. Accordingly,

generalized additive mixed models (GAMM) were chosen to incorporate the non-linear form and to include an autoregressive term.

The R package 'mgcv' (Wood 2017) was used for GAMM model evaluation. Models were built from the logical predictors using (1) time, (2) month number, (3) Dipole Mode Index [a continuous variable], and (4) the Nino 3.4 index [continuous variable]. Covariates were modelled as fixed effects (i.e. affecting all the data). Non-significant terms were removed from models.

All analyses were conducted in R using the R script file "*Phuket CRW data – GAMM models.R"* which contains further details and notes on use. Values for monthly DMI and Nino 3.4 obtained from the earlier analyses of these indices had been stored in the Excel file "*DMI Nino 3.4 monthly to 14 Mar 2020.csv*" and were input to the R code were indicated.

Hotspot daily data were meaned over monthly periods so that these could be aligned to DMI and Nino 3.4 monthly data for the analyses.

<u>Data Exploration</u>

Data exploration was applied following the protocols in Zuur et al. (2010). The presence of possible outliers was investigated using Cleveland dotplots or boxplots.

Hotspot data plotted by month (Fig S5.2) reveals a seasonal pattern with higher Hotspots in April/May each year. This is consistent with our observations that the highest sea temperatures occur at this time of year at the end of the dry season. Negative outliers occurred in 1996 and 2000. A positive outlier was present in 1988. The outliers were retained in the dataset as valid readings.

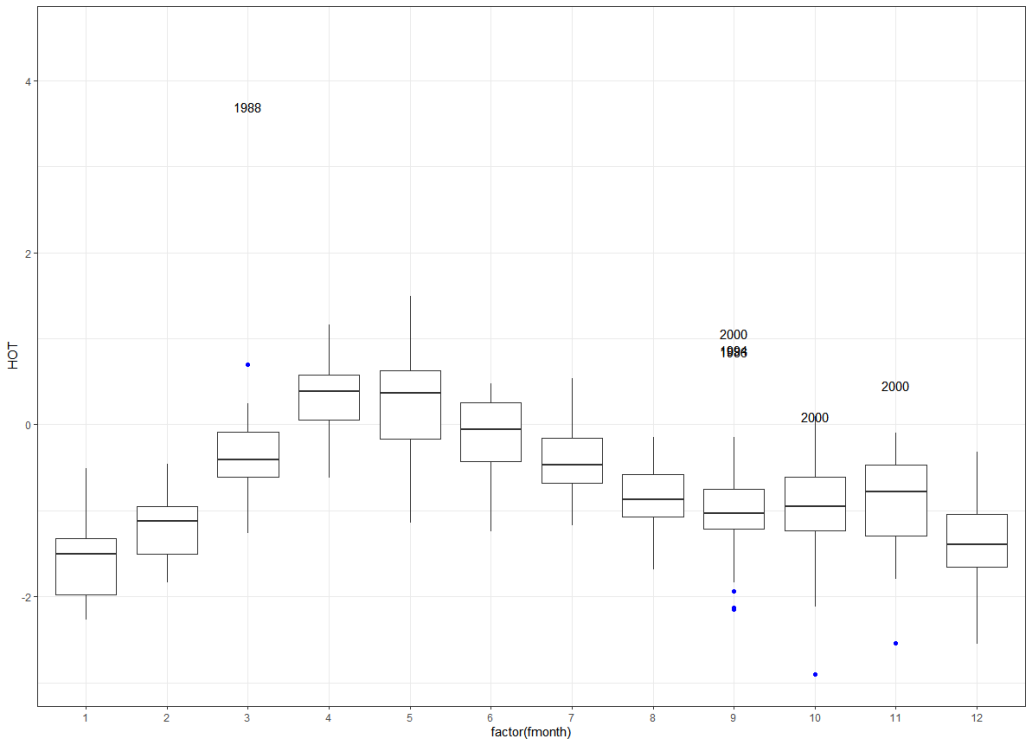

Fig S5.2. Boxplot of monthly Hotspot data (1985-2020) for the grid square, 7.825°N 98.475°E, ~7.8km to the east of Cape Panwa, Phuket. Outliers, in blue, are identified by the year in which they occur. Outliers are 1.5 times the inter quartile range above (3rd Quartile) or below (1st Quartile).

The observed pattern of seasonality confirmed that a term (month) for this should be included in the model.

Collinearity between the potential covariates was assessed using multi-panel scatterplots, Pearson correlation coefficients and variance inflation factors (VIF). VIFs were computed using the 'corvif' function (Zuur et al. 2009).

There was no collinearity apparent between the chosen covariates (DMI, month, time, and Nino 3.4) (Fig S5.3). This was confirmed by variance inflation factors (VIF) which were approximately 1.0 in each case. A cut off value of 3 for the VIF was used.

Covariate	GVIF
DMI	1.1321
month	1.0059
time	1.0254
Nino3.4	1.1022

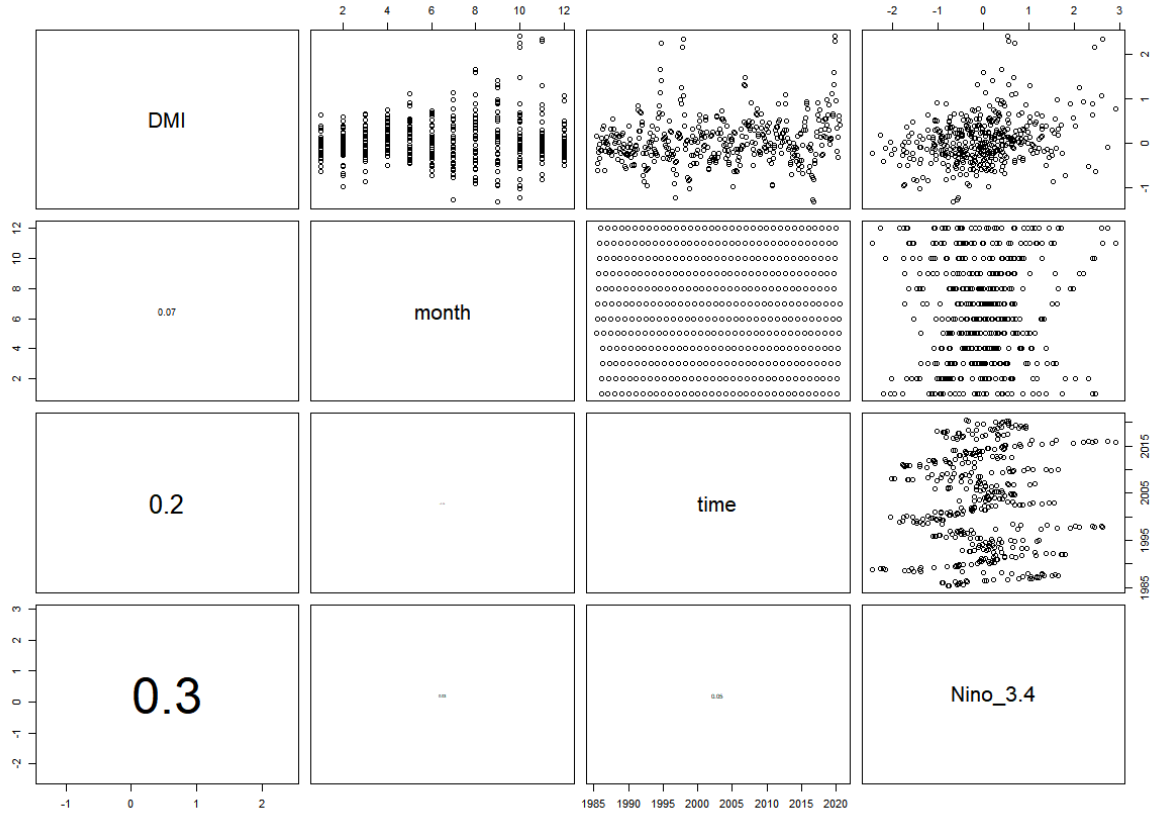

Fig S5.3. Pair plots for covariates (top right) with Pearson correlation coefficients (bottom left).

Datasets were checked for missing data. The relationships between Hotspot and individual covariates were visualised using multi-panel scatterplots (Fig S5.4).

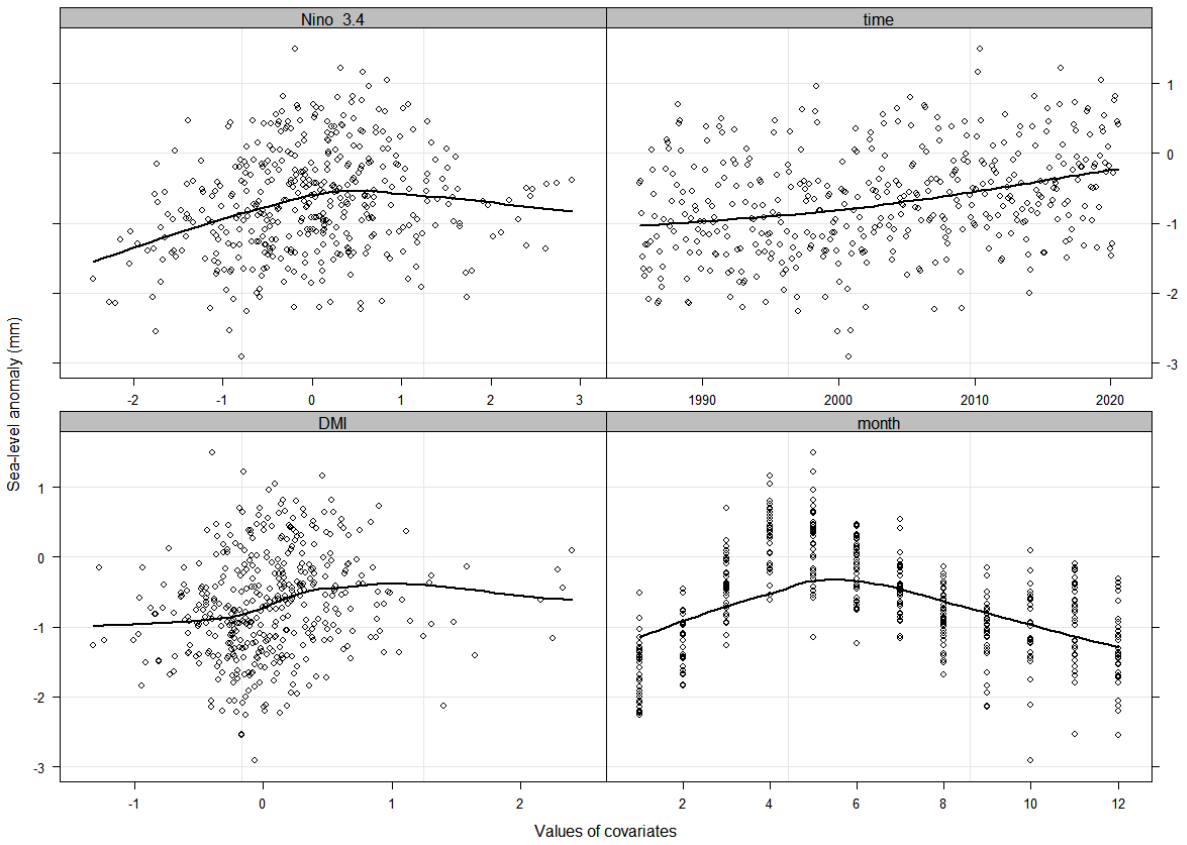

Fig S5.4. Covariates plotted against Hotspot. Monthly data. A Loess smoother has been added to each plot to aid interpretation.

Model choice and selection

Exploratory ACF/PACF plots identified autocorrelation in the Hotspot data and were used to indicate the autoregressive AR() term which would probably be needed to remove this. A first order AR(1) term was found to be suitable. For this term, month was nested within year (i.e. the covariate determining the ordering of the residuals was 'month' and the correlation applied within measurements made in one year).

The smoothing method REML (restricted maximum likelihood) was chosen since this is known to give stable, reliable results. 'mgcv' was initially allowed to run in its automated mode to select appropriate smoothing parameters and basis functions. If necessary specific values for these parameters were later added to the model.

Thin plate spline (tp) basis functions were used for time and DMI variables. The seasonal term (month) was initially modelled using cubic cycle (cc) for the repeating seasonality but where patterns still remained in the residuals, month was modelled as a factor to allow different intercepts for each month.

An interaction between time and seasonality (i.e., the monthly seasonality changing over the timescale of the times series) was investigated by the inclusion of an interaction term using a tensor product interaction to account for the different scales of the time and seasonal components. The interaction term was however non-significant and was dropped from further models.

Several GAMM models, all of which included terms for Dipole Mode Index (DMI), Nino 3.4, time, and seasonality (month) were computed. In the 'M2' and 'M2aa' model DMI and Nino 3.4 were included as smoothing terms but Nino 3.4 had an effective degrees of freedom (edf) close to 1, indicating a linear term. The term for DMI was not significant and was removed in the 'M2_no_DMI' model.

Model selection was based upon AICc (AIC with small sample bias adjustment). Adjusted R^2 values were used to assess the amount of variation in the sea-level time series accounted for by the model and individually for each additive term.

A range of models were compared and the best fit chosen based upon lowest AICc using the package 'MuMIn' and the 'model.sel' function (Barton 2020).

	X	family	df	logLik	AICc	delta	weight
M2_no_DMI$lme	+	gaussian(identity)	8	-146.75	309.84	0.000	0.974
M2$lme	+	gaussian(identity)	10	-148.29	317.12	7.275	0.026
M2aa$lme	+	gaussian(identity)	20	-149.78	341.64	31.797	0.000
M2x$lme	+	gaussian(identity)	31	-140.79	348.65	38.806	0.000

The optimum model (M2_no_DMI) appears first in the list and explained 73.2% of the variation in Hotspot. The model included fixed effects of Nino3.4, time, and seasonality (month) all as smoothing terms.

R code for the model fit was:

```
M2_no_DMI <- gamm(HOT ~ s(time) + s(month, bs = "cc") + s(Nino_3.4), data = monthly_Hotspot_DMI, correlation = corAR1(form = ~ month | year), method="REML", select=FALSE, knots=list(month=c(0,12)), control = lmeControl(opt = 'optim', msVerbose = TRUE))
```

All model terms were significant.

```
Family: gaussian
Link function: identity

Formula:
HOT ~ s(time) + s(month, bs = "cc") + s(Nino_3.4)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.69313    0.03319  -20.88   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
             edf Ref.df     F  p-value
s(time)    1.130  1.130 41.87  < 2e-16 ***
s(month)   7.555  8.000 87.05  < 2e-16 ***
s(Nino_3.4) 1.009  1.009 17.24 3.91e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.732
  Scale est. = 0.15099   n = 423
```

Phi=0.5658

<u>Model checking</u>

Models were checked using the 'mgcv' function 'gam.check' for convergence, and the appropriate number of basis functions. The 'concurvity' function was used to check for correlation between variables. Model validation also included the following steps:

1. To verify homogeneity of variance, residuals were plotted versus the fitted values.
2. To verify model misfit (or independence), residuals were plotted versus each covariate in the model.
3. To verify independence (repeated measurements over time), auto-correlation plots were created from the residuals.
4. To verify the normality assumption, a histogram of the residuals was created.
5. The model was inspected for influential observations.

```
> gam.check(M2_no_DMI$gam)

'gamm' based fit - care required with interpretation.
checks based on working residuals may be misleading.
Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

             k'  edf k-index p-value
s(time)    9.00 1.13    0.46  <2e-16 ***
s(month)   8.00 7.55    1.04    0.78
s(Nino_3.4) 9.00 1.01    0.99    0.36
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residuals were normal, homoscedastic and the AR(1) term removed the autocorrelation (Fig S5.5). There was no problematic concurvity (worst case 0.42).

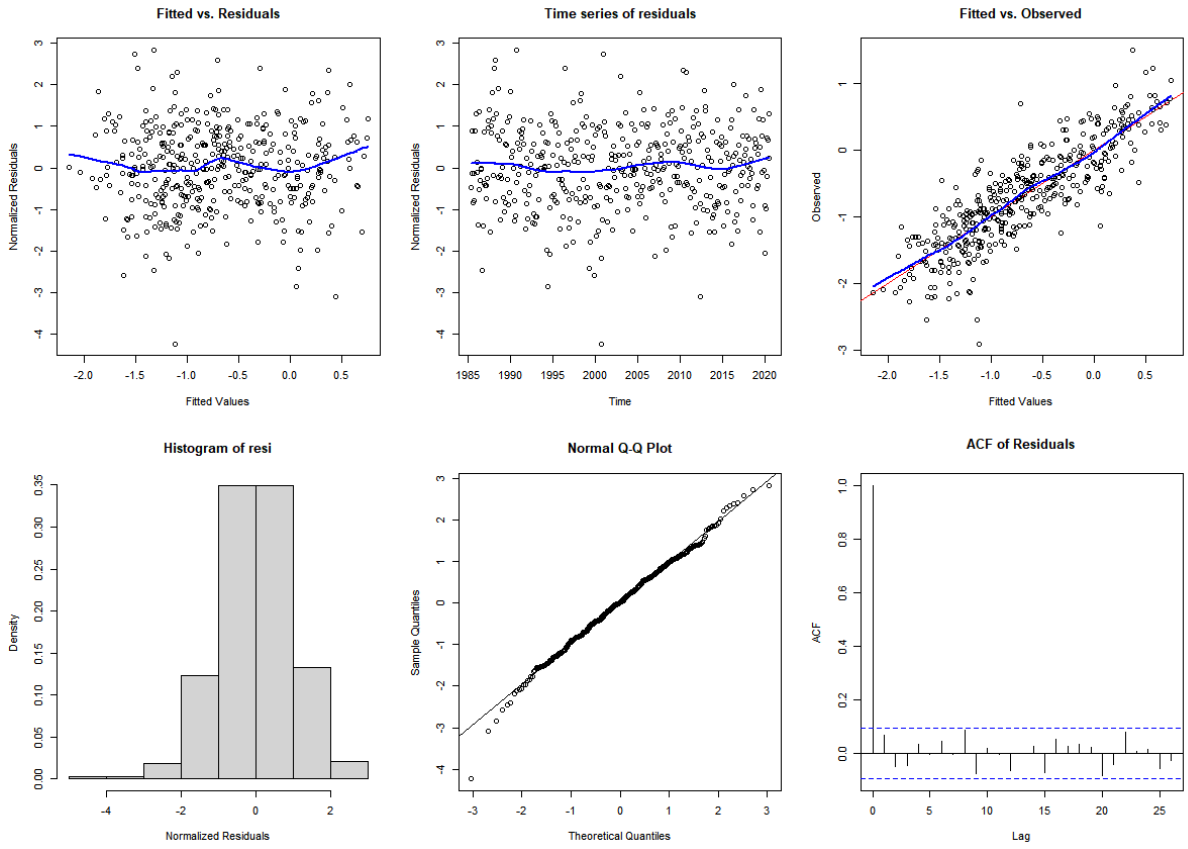

Fig S5.5. Diagnostic plots for the final model.

Plots of the model residuals against each covariate (Fig S5.6) indicated that there were no patterns remaining in the residuals.

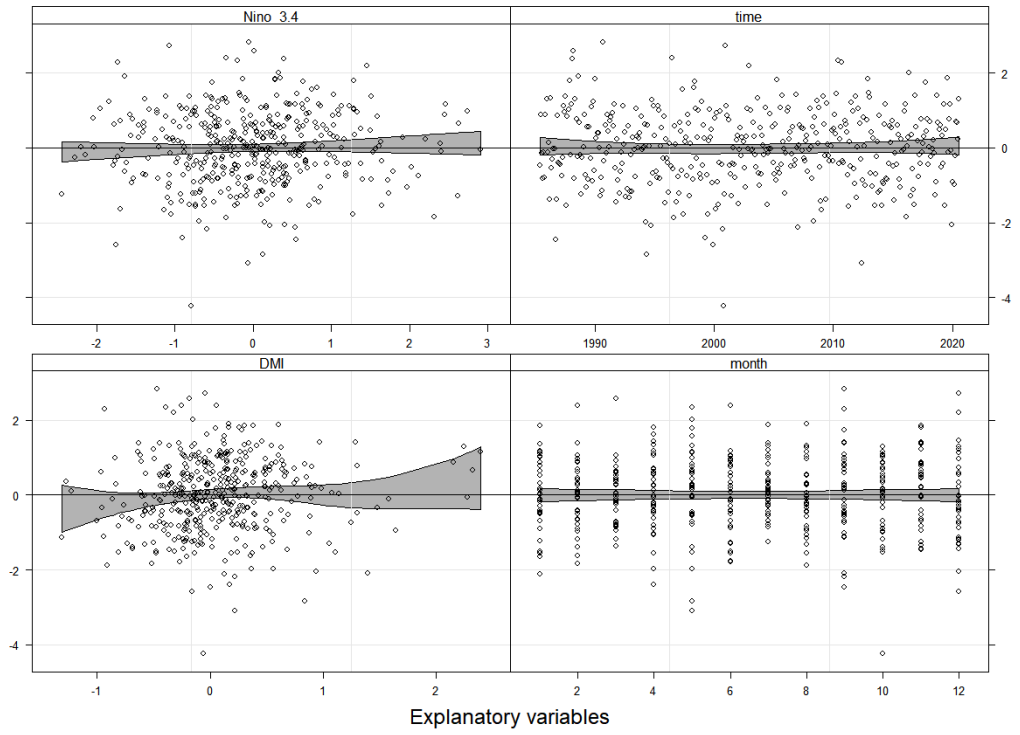

Fig S5.6. Plots of residuals versus covariates (DMI included although not in the model). Smoother and 95% point-wise confidence intervals (grey shade). The confidence intervals contain 0 for all covariate values indicating no significant patterns.

Finally, a conditional boxplot of the residuals by month (Fig S5.7) confirmed no significant residual month effect.

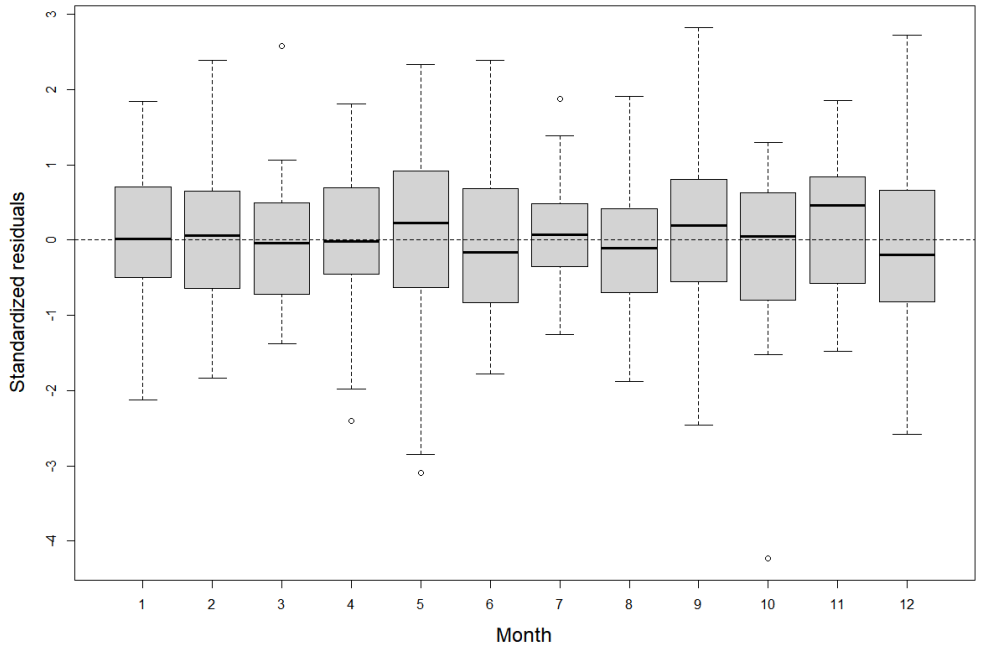

Fig S5.7. Conditional boxplots of residuals grouped by month.

Visualising the model output

As well as inspecting model output in the form of partial plots, the overall model and individual terms were extracted and plotted using the 'predict' function in 'mgcv'.

To represent the output from the model the effect of each term was examined separately as well as the overall model. Since the model is additive, the overall model is a summation of each term (time, Nino 3.4, month) together with the intercept.

Partial effect plots for each model term (Fig S5.8) show how each term separately contributes to the model whilst the other covariates are held at their median value. The relationship with Nino 3.4 was effectively linear and positive, as might be expected.

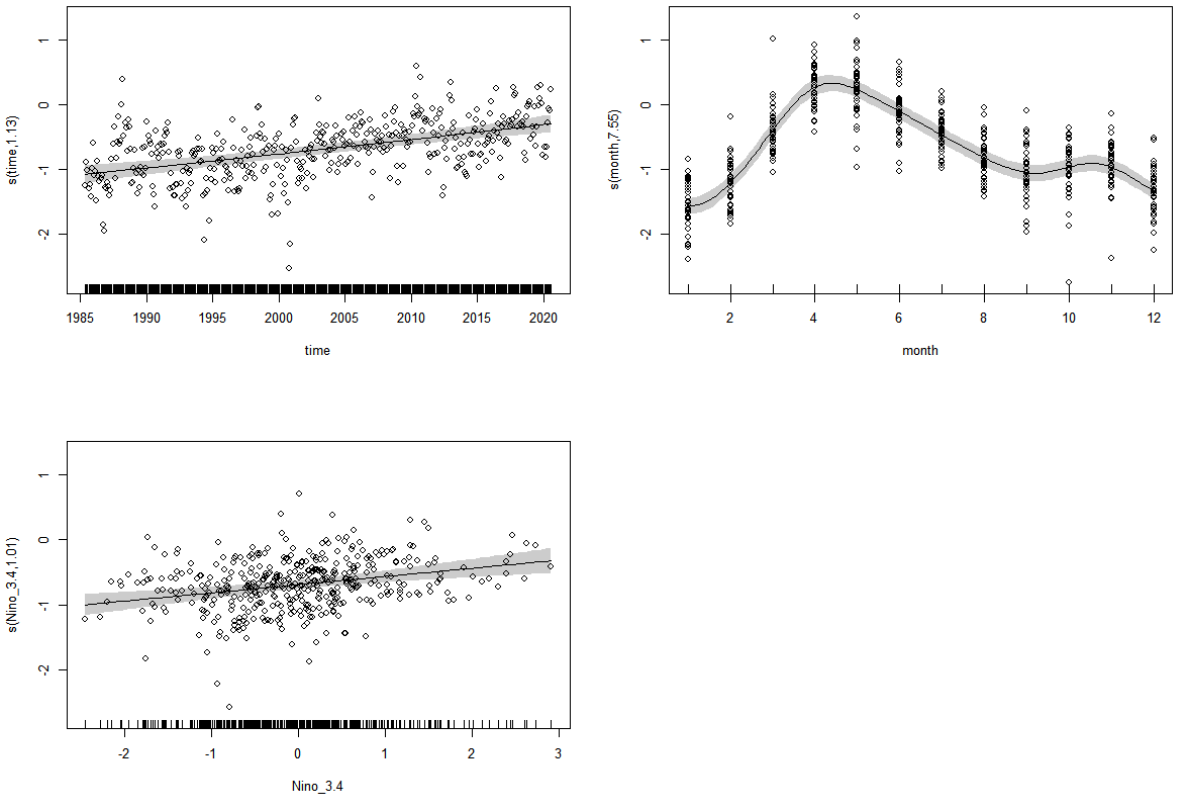

Fig S5.8. Partial effect plots for the GAMM model. (1) Nearly linear relationship between time and Hotspot (2) Non-linear relationship between Month and Hotspot (3) Nearly linear relationship between Nino 3.4 and Hotspot. Residuals plotted as open circles. Partial effects include model intercept. Shaded areas/ dotted lines are 95% CI.

Individual Model Terms

The overall model fit for all terms is shown in Fig S5.9. Separate fits for each of the terms are shown in Figs S5.10-5.14. The model explained 73.2% of the variation in the data with the seasonality (month) accounting for ~61.5%, time ~9.6%, and Nino3.4 ~3.7%.

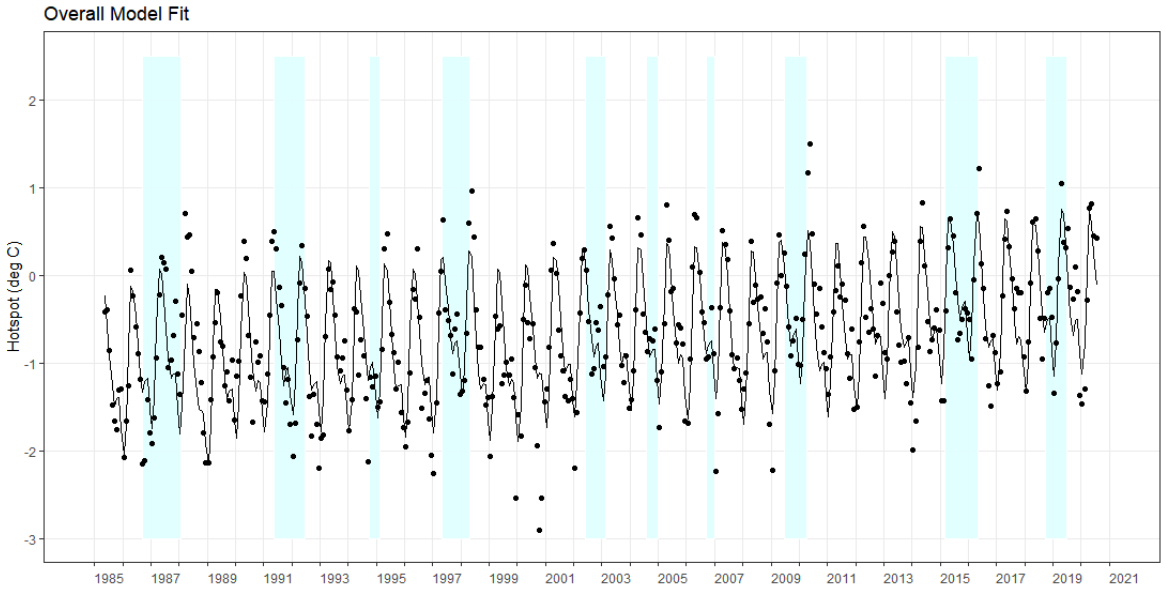

Fig S5.9. Monthly mean Hotspot data (black dots) with overall model fit (black line). Cyan shaded time periods represent timing of El Niño (Nino3.4 >0.5 ºC).

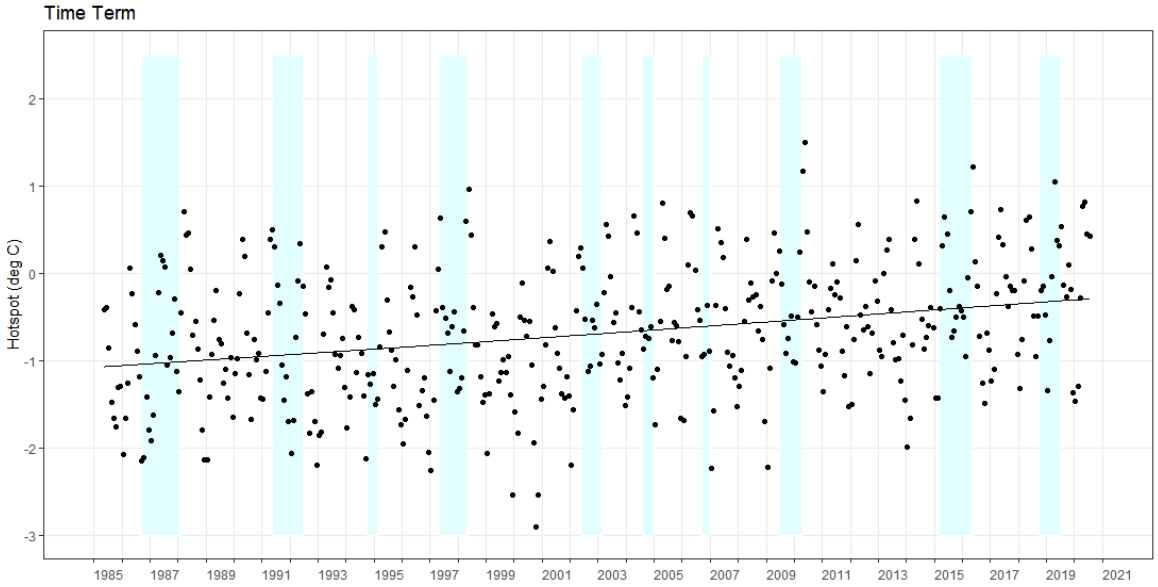

Fig S5.10. Monthly mean Hotspot data (black dots) with time term fit (black line). Cyan shaded time periods represent timing of El Niño (Nino3.4 >0.5 ºC).

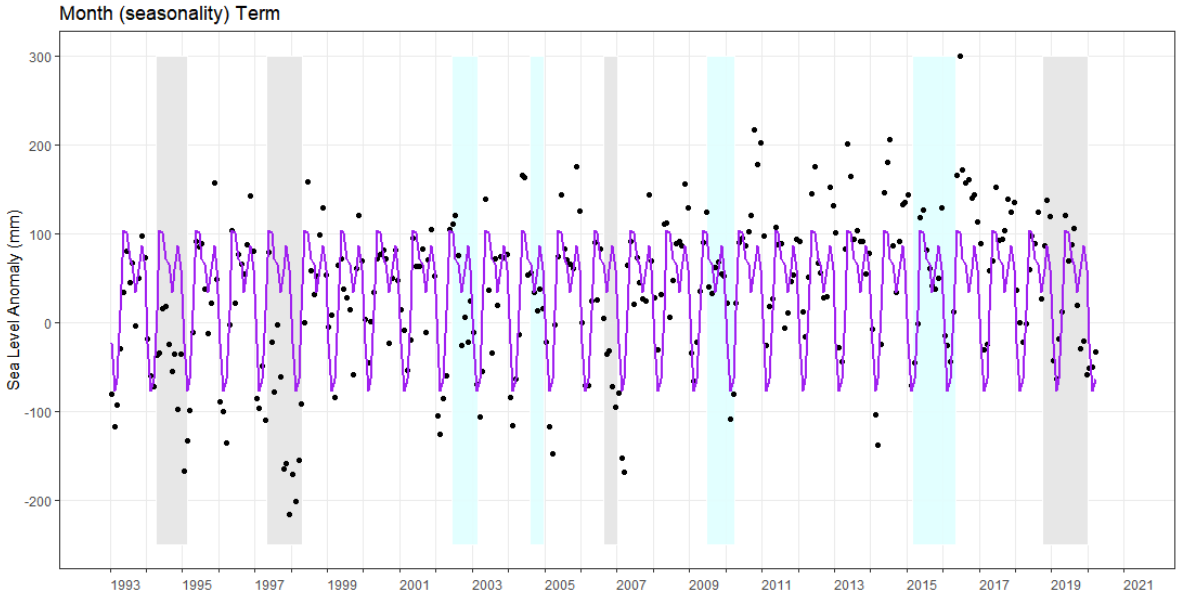

Fig S5.11. Monthly mean Hotspot data (black dots) with Month term fit (purple line). Cyan shaded time periods represent timing of El Nino (Niño3.4 >0.5 °C).

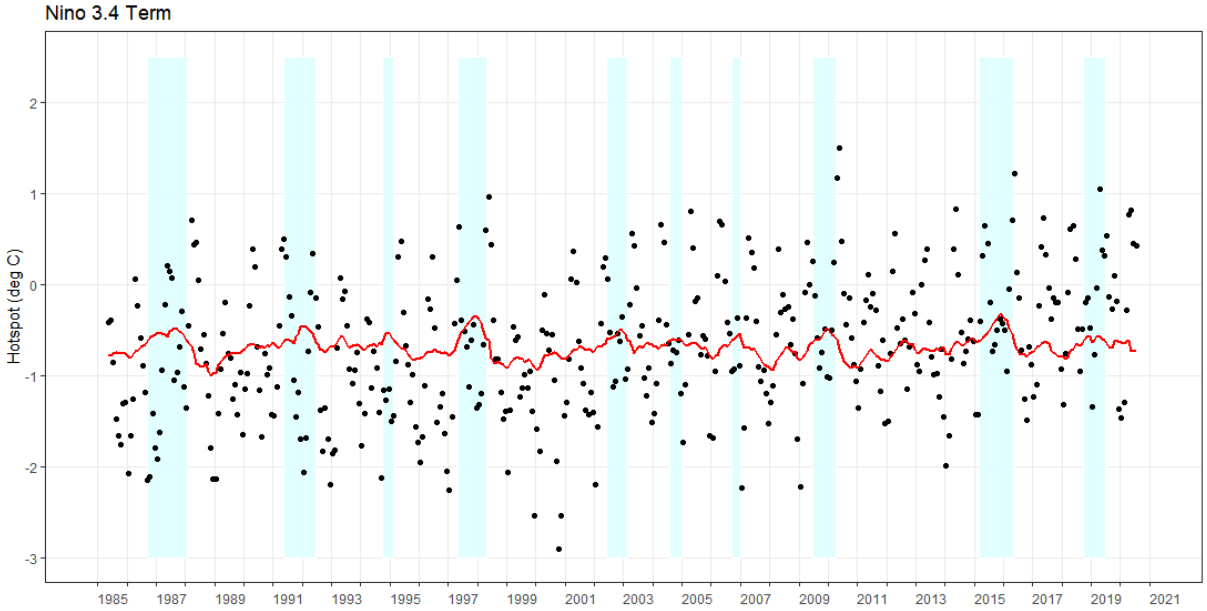

Fig S5.12. Monthly mean Hotspot data (black dots) with Nino 3.4 term fit (red line). Cyan shaded time periods represent timing of El Niño (Nino3.4 >0.5 °C).

The fit for Nino 3.4 (Fig S5.12) shows that this term produces a small peak for each El Niño with the exception of 2019.

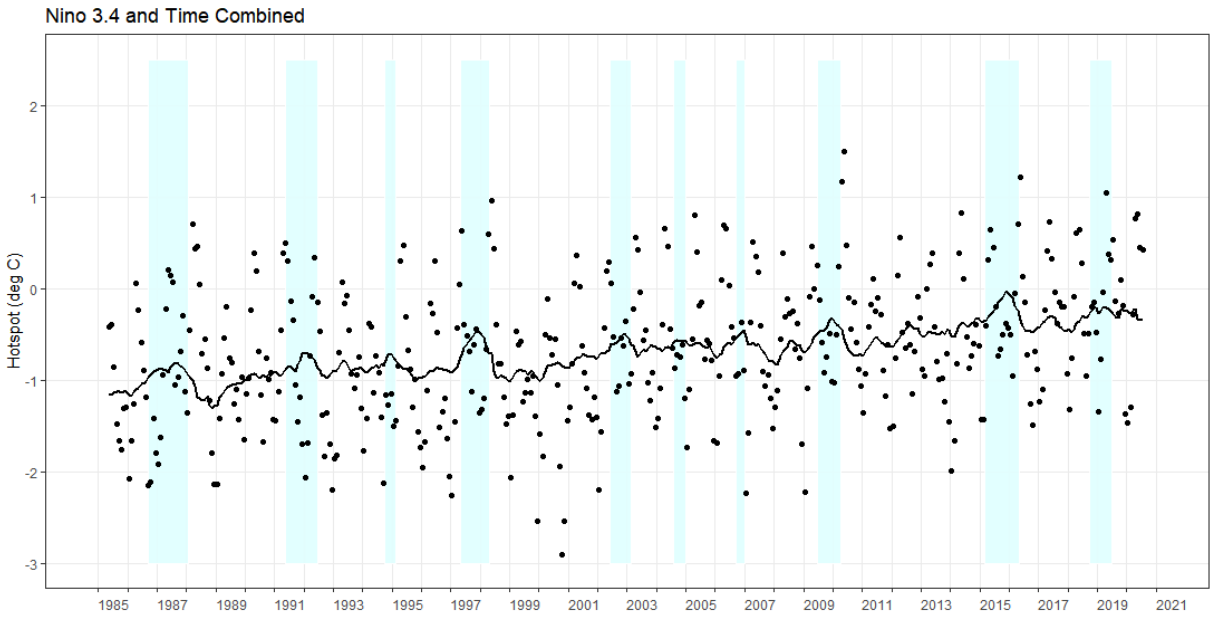

Fig S5.13. Monthly mean Hotspot data (black dots) with combined Nino 3.4 and time terms fit (red line). Cyan shaded time periods represent timing of El Niño (Nino3.4 >0.5 ^{o}C).

The plot of the overall model with seasonality removed (Fig S5.13) shows that the increase in Hotspot temperatures over time raises the magnitude of the relative peaks during El Niño events.

6. GAMM modelling patterns of the phototransect coral cover time series.

R code "*Phototransect Analysis in R FINAL.R*"

Tide Gauge relative sea level data

Tide Gauge data was obtained for the Royal Thai Navy tide station at Ko Taphao Noi (source Royal Thai Navy 1977-Dec 2020 as monthly means). The RTN dataset had 2 missing monthly means (Oct 2001 and July 2002). Values for these months were imputed using the means of the respective monthly values for the preceding 10-year period (1991-2001).

Data was converted to monthly anomalies using the long term mean from 1 Jan 1977 to 1 May 2020. A correction was also applied for changes in the tide gauge datums during the periods 1 Nov 2012 – 1 Jan 2013 (+22mm), 1 Jan 2013 – 1 Jan 2015 (-61mm) and 1 Jan 2015 onwards (-66mm) (for further details of the necessity for this please contact the authors). Finally, running means of the anomalies were computed for periods of between 1 and 14 months prior to the respective monthly mean value.

Degree Heating Week data (DHW)

Hotspot data from the Coral Reef Watch 5km satellite coral bleaching heat stress monitoring products (Version 3.1) (Liu et al. 2014) were downloaded for a 5km grid square 7.825^{o}N 98.475^{o}E, 7.8 km to the east of the study site at Cape Panwa, Phuket covering the time period 1 July 1987 to 13 January 2021. Data are daily values of Hotspot, which is the amount by which the daily mean satellite sea surface temperature is above the highest climatological monthly mean in ^{o}C. A corresponding value for DHW (a measure of accumulated heat stress in ^{o}C-weeks) was computed from the Hotspot data using the same methodology as Coral Reef Watch but over a range of different time periods (30 days to 360 days in 30-day increments). The examination of extended DHW time periods was based on earlier findings that mortality of corals on the reef flat when exposed to elevated sea surface temperatures can take between 15-44 weeks to appear and thus be reflected in measurements of reduced coral cover (Brown and Phongsuwan 2012).

The processed data for SLA and DHW is stored in the Excel File "*Coral Cover and Covariates.csv*" together with the data for coral cover on each phototransect.

R code for the analysis is in file "*Phototransect Analysis – GAMM models.R*".

Data Exploration

Data exploration was applied following the protocols in Zuur et al. (2010).

Visualising the Data

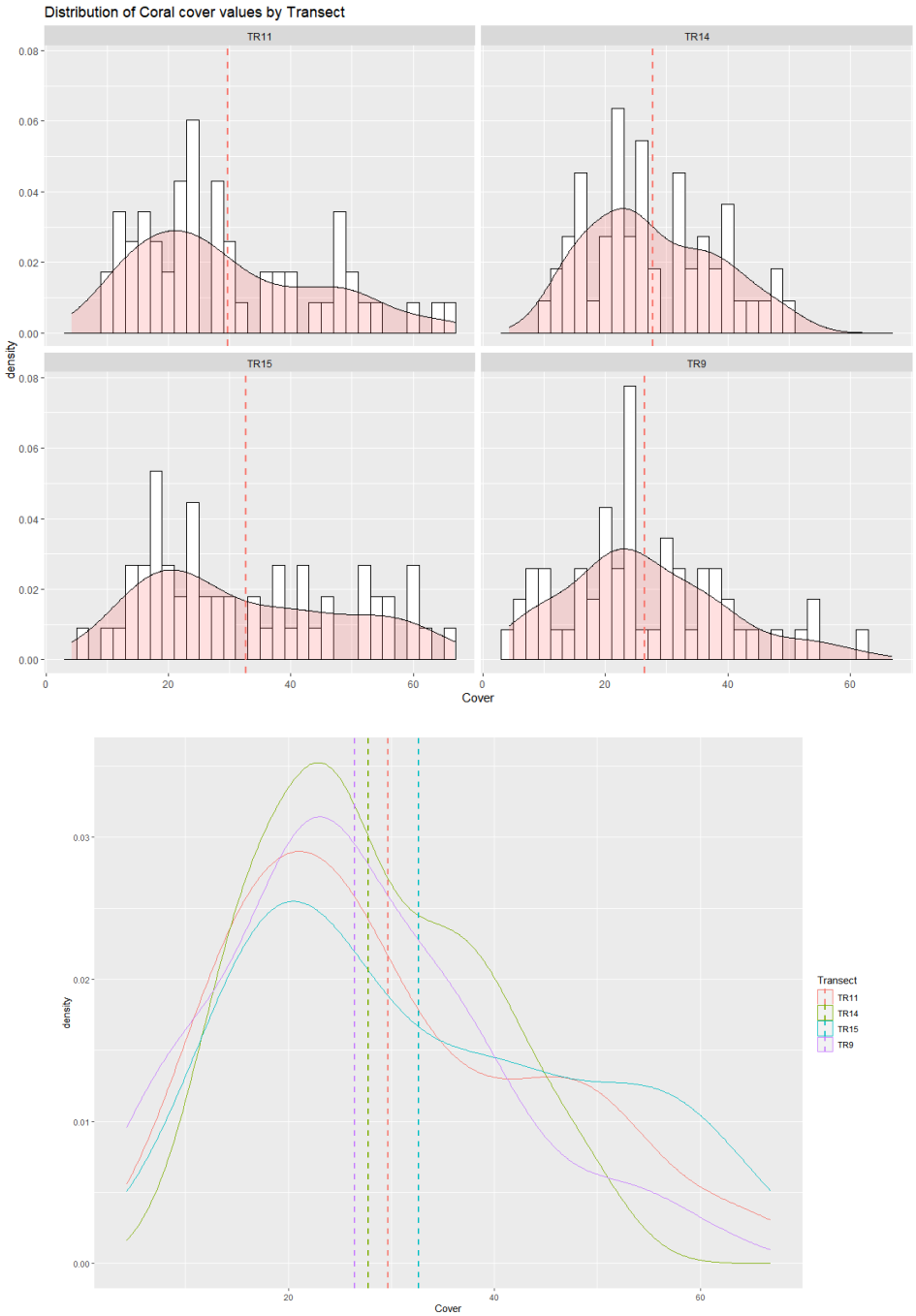

Fig S6.1. Coral Cover - Histogram/Density Plots for each Transect separately (top) and Density Plots for all Transects (below). Pink shading in the top plot is the corresponding density plot. Vertical dotted lines are the means.

Histogram and density plots (Fig S6.1) show that the distribution of coral cover data differed between transects. Although we see from these plots that the distributions are not especially 'normal', when using a model in GAM or GAMM it is not the normality of the data we are concerned about but the residuals from the model.

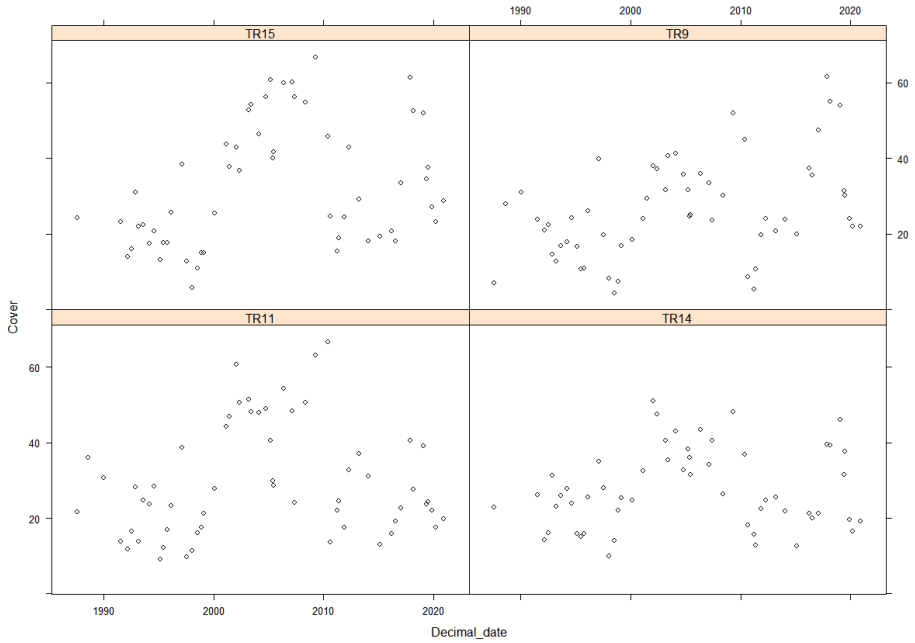

Fig S6.2. Data spread for each transect.

If the spread of data for each transect appears to be very different (Fig S6.2) then we might have to consider a weighting for each transect to account for heterogeneity such as a model with a term "weights=varIdent(form=~1|fTransect)". The weights option here implements a variance structure to deal with heterogeneity where each Transect group is allowed a different variance (see Zuur et al. (2009)).

These plots do not reveal any major differences in the spread of the data for each Transect so weighting is unlikely to be required.

Choice of Covariates

It was hypothesised that coral cover would depend on the covariates (1) time, (2) SLA, (3) DHW, (4) seasonality (month) and from the above visualisation a term with transect as a factor to allow different intercepts for each transect.

SLA

The choice of which time period for the SLA term was based upon earlier findings by (Brown et al. 2011) who showed that coral cover and sea level were correlated on the basis of the running mean sea level over a period of several preceding months, with a significant correlation for time periods of 6-16 months, together with visualisation of the data (Fig S6.3) which showed a consistent relationship between cover and SLA had evolved by the 6 month point. Accordingly a SLA term for the 6 month running mean was chosen (X0.6.1) for inclusion in the model.

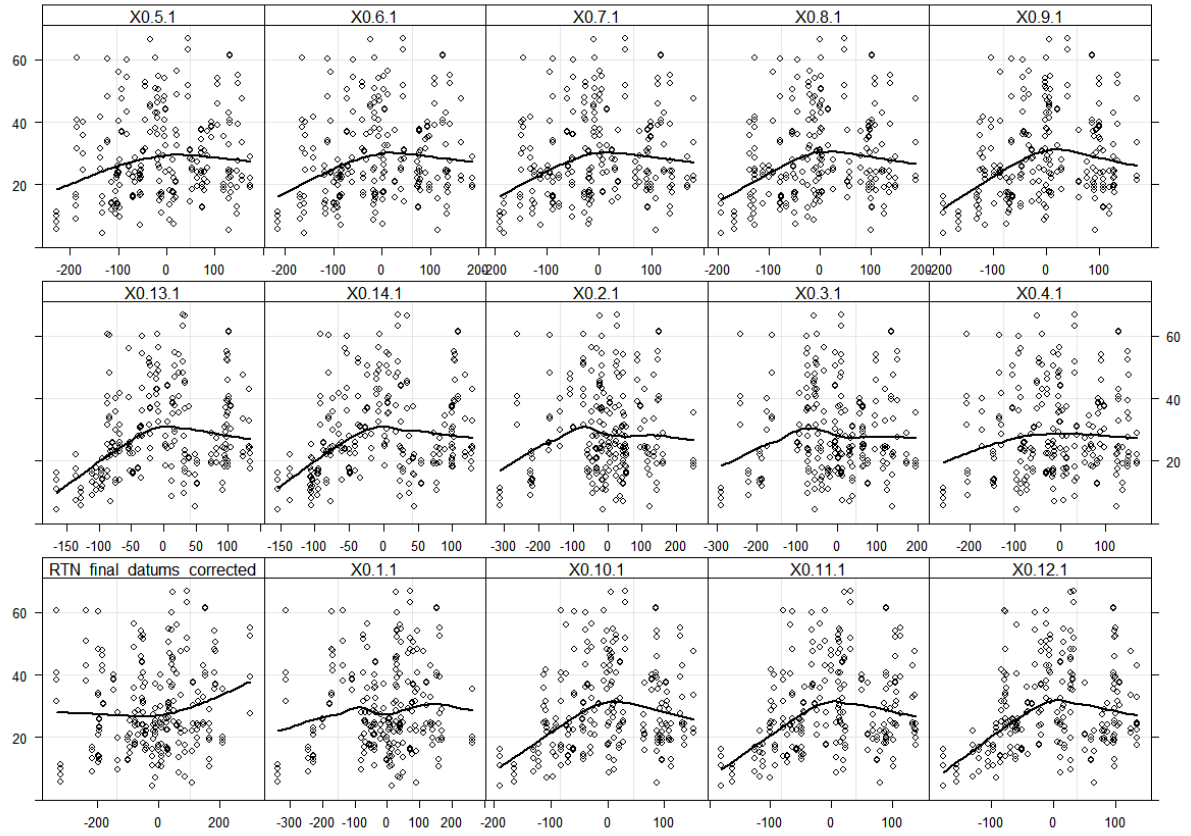

Fig S6.3. Relationship between coral cover and tide gauge (corrected) values of mean sea level (SLA) computed as the running mean for time periods of between 1 and 14 months. 'RTN final datums corrected' is the mean sea level for the month of sampling. A Loess smoother has been added to each plot to aid interpretation.

DHW

The use of DHW as a measure of accumulated heat stress (°C-weeks) for assessing coral bleaching and mortality is an established technique (Liu et al. 2014). The Coral Reef Watch measure of DHW is computed for a period of 84 days before the date in question. This and our computed values for DHW ranging from 90 days to 360 days are shown in Fig S6.4 plotted against coral cover. DHW values >4°C-weeks are known to accompany significant coral bleaching and those >8°C-weeks are associated with severe, widespread bleaching and significant mortality (Liu et al. 2014).

The spacing of the sampling times for coral cover will in part determine our choice for the period over which to compute DHW for use in the model. This is because the occurrence of a DHW impact on coral mortality and therefore cover is a random event linked in part to El Niño or *p*IOD events (see for example Main Paper Figure 8). The irregular coral cover sampling frequency (maximum 18 months, minimum 1 month, mean 7 months) requires a suitable length of DHW computation in order to capture periods of heat stress which may have and continue to impact the coral cover on the sampling date. From Fig S6.4 we can deduce a number of features. Firstly, that the extreme values of DHW (11.5°C-weeks) are likely to have a larger influence on the overall relationship but this is nonetheless as expected. Secondly, that a large proportion of the DHW data equals 0 (zero). Again this is entirely expected because by definition the DHW metric only captures Hotspot values which are >1°C above the maximum monthly mean, thus the majority of values will =0.

Although all the plots show a downward trend of coral cover with increasing DHW, as we might expect, the shorter time periods also display values of DHW <2°C-weeks correlating with a dip in cover. We can also see that only when we reach DHW 120 and above do we capture the highest value of DHW (11.5°C-weeks) which occurred in June 2010 prior to the cover sampling date in August when mean cover had reduced to 16.3%. This indicates how too short a DHW window will distort the picture. Thus in order to adequately capture the possible impact of DHW on the 10 instances in the data where the sampling interval was 12 months or greater, a DHW period of 360 days was selected (DHW_360) as the model term.

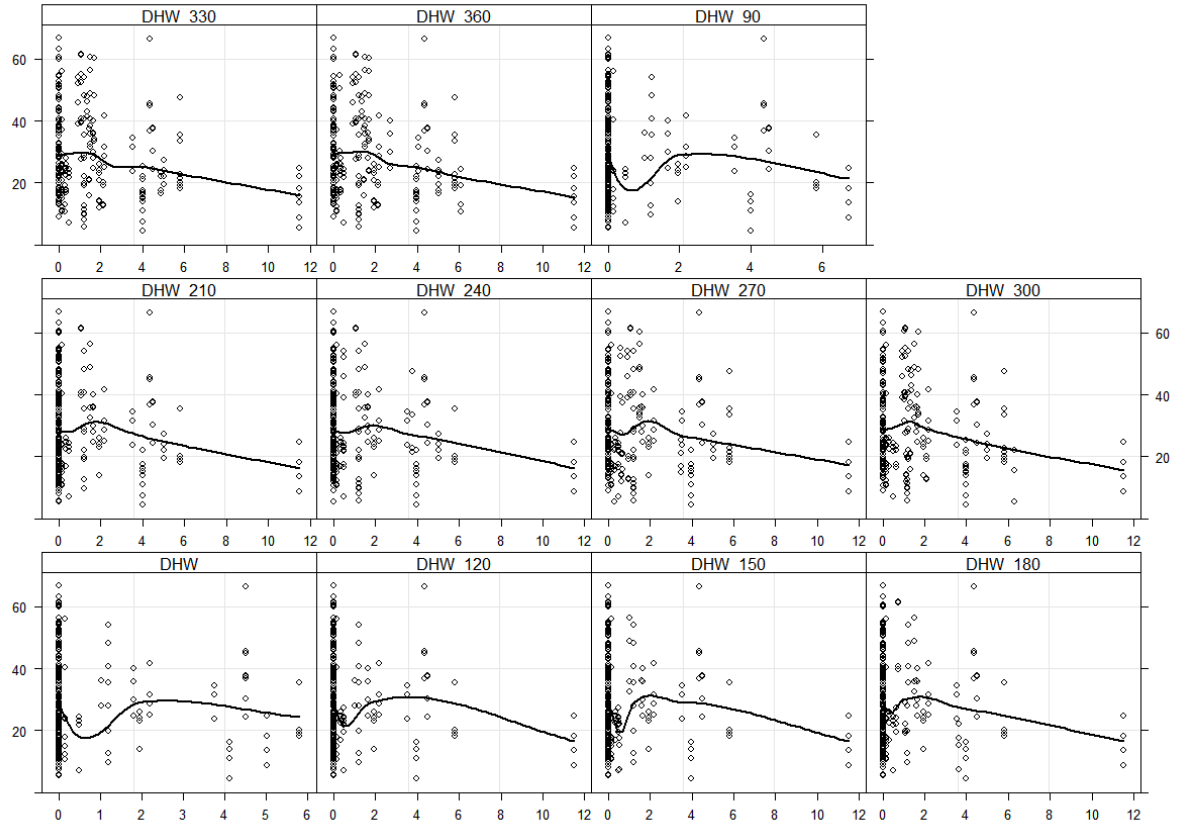

Fig S6.4. Relationship between coral cover and DHW (the Coral Reef Watch metric) and computed values of DHW for between 30 and 360 days. A Loess smoother has been added to each plot to aid interpretation.

Identifying collinearity between covariates

Using the selected covariates, collinearity was computed and plotted between each covariate (Fig S6.5).

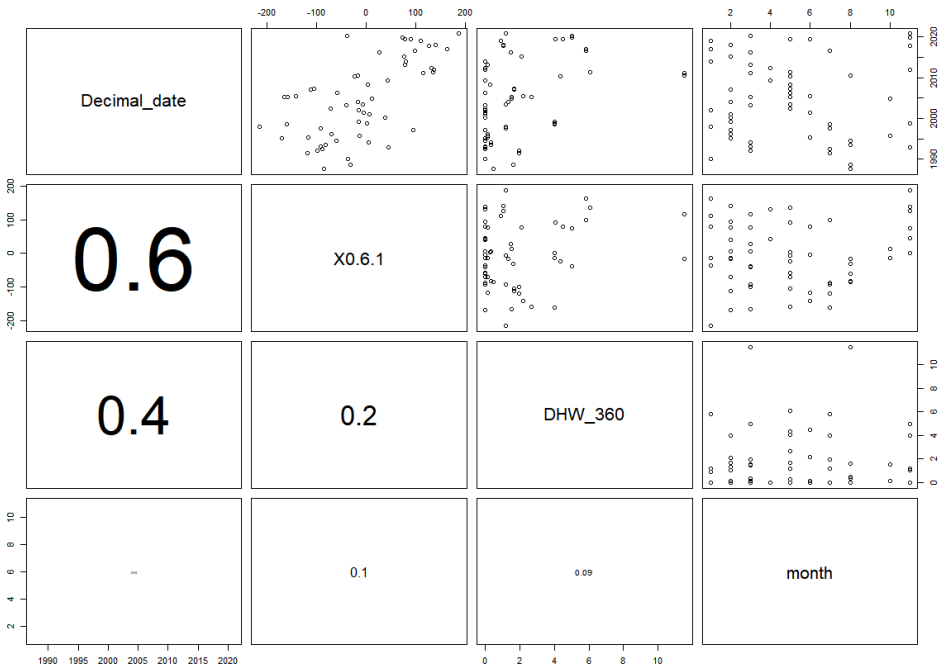

Fig S6.5. Collinearity between covariates.

There are correlations between Date (Decimal_date) and Sea Level (X0.6.1) and to a lesser extent between Date and DHW as shown in Fig S6.5. A correlation of 0.8 is considered critical, whereas values between 0.5-0.7 indicate that care is required when correlated covariates are included in the model. A further way to check for collinearity is by VIF.

Variance Inflation Factors (VIF)

VIF(j) is the factor by which the variance of βˆj is increased over what it would be if xj was uncorrelated with the other independent variables. If all values of VIF(j) are near 1, then collinearity is not a problem and there is no correlation between this covariate and any others. VIFs between 1 and 5 suggest that there is a moderate correlation, but it is not severe enough to warrant corrective measures. VIFs >5 represent critical levels of multicollinearity where the coefficients are poorly estimated, and the p-values become questionable. VIF(j) >10 indicates serious collinearity (Ieno and Zuur 2015).

	GVIF
Decimal_date	**2.042**
X0.6.1	**1.795**
DHW_360	**1.212**
Month	**1.047**

In this case the VIF values indicate that including all these covariates in the model should not prove problematical.

Initial look at relationships between covariates and Coral Cover

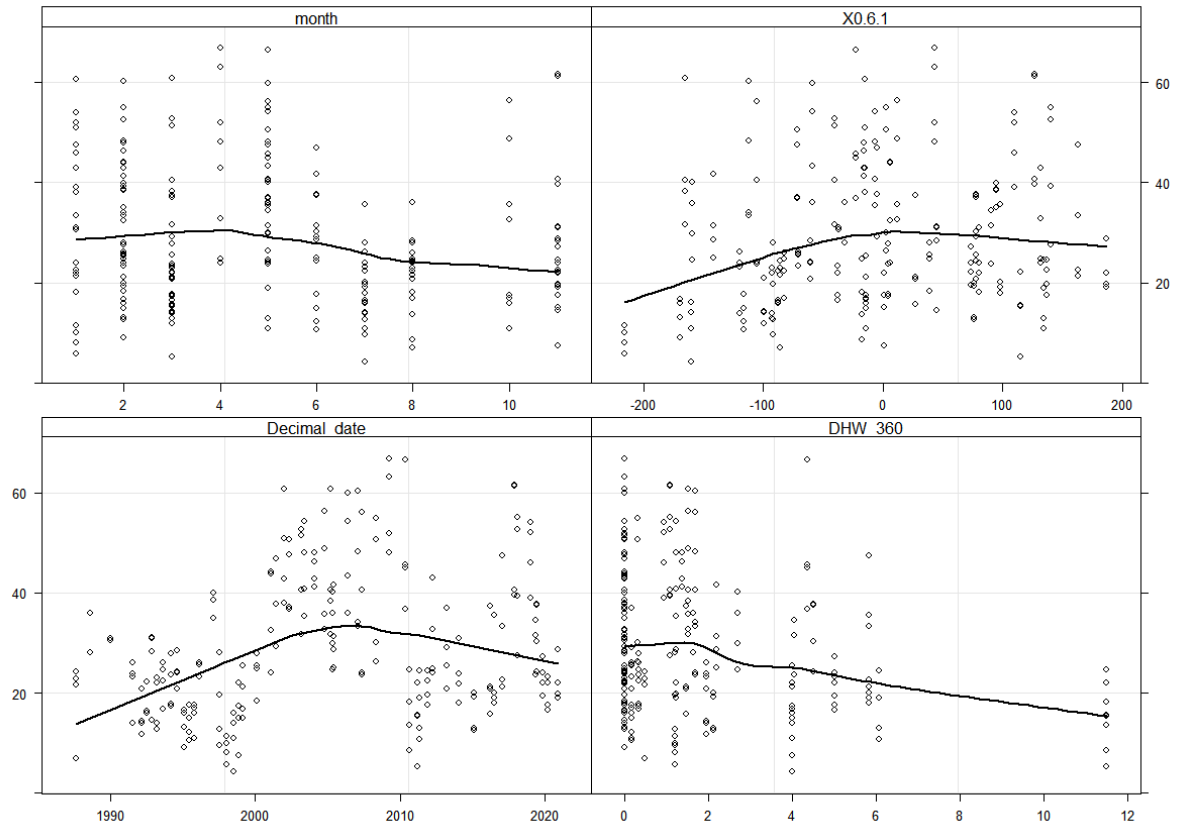

Fig S6.6. Covariates and their relationship with coral cover.

Fig S6.6 shows that coral cover is depressed when sea-level anomalies are negative, whilst positive anomalies appear not have little relationship with cover (depending on the confidence intervals of these plots). For DHW >2°C-weeks we see a steady drop in cover which is what we might expect. On a time basis cover appears to have increased from 1987 to about 2005, decreasing thereafter. For the month of coral cover sampling there appears to be an effect whereby sampling dates in the boreal autumn/winter produce lower cover values.

7. Modelling patterns in the coral cover time series.

An exploratory Generalised Additive Model (GAM) was first run and the autocorrelation (ACF) function used to check for autocorrelation which was found to be present in the cover data over time, necessitating a GAMM with an AR() term.

```
Full_model <- uGamm(Cover ~
                    s(Decimal_date, k=25) +          # Time
                    s(SLA, k=25) +                   # sea level
                    fTransect +                      # Transect number as a categorical term
                    s(month) +                       # month
                    s(DHW, k=15),                    # Degree Heating Weeks
                    data = monthly_subset,
                    correlation = corAR1(form = ~ 1 | year), method="REML")
```

The GAMM was constructed using time, sea-level anomaly (SLA – X0.6.1), seasonality (month), and Degree Heating Weeks (DHW_360) as smoothed terms and the individual transect as a categorical term. The autocorrelation present was modelled using the correlation structure "**form = ~1|year**" such

that the temporal order of the data is specified by the variable "year" and the same residual correlation structure is assumed each year - shown by only one value for phi in the $lme component of the model. It is acknowledged that this structure does not consider the potential for residual variation from year to year.

Including the **fTransect** categorical covariate means that the model uses one intercept and one smooth curve for each smoothed covariate and separate regression parameters for each of the 4 transects. Each of these regression parameters is used as a correction of the intercept for a particular plot. As a result the full model predicted values consist of 4 parallel smooth lines. The resultant model thus contains an additional 4 regression parameters and the model allows only for statements and predictions applying to these 4 Transects not for transects in general.

Model selection table

	(Int)	fTr	s(Dcm_dat,25)	s(DHW,15)	s(mnt)	s(SLA,25)	df	logLik	AICc	delta	weight
32	29.85	+	+	+	+	+	14	-804.448	1638.9	0.00	0.568
24	29.85	+	+	+		+	12	-806.983	1639.4	0.55	0.432
8	29.87	+	+	+			10	-817.352	1655.7	16.85	0.000
16	29.88	+	+	+	+		12	-816.495	1658.4	19.57	0.000
31	29.17		+	+	+	+	11	-818.033	1659.3	20.42	0.000
23	29.17		+	+		+	9	-820.706	1660.2	21.37	0.000
7	29.18		+	+			7	-830.073	1674.7	35.78	0.000
15	29.18		+	+	+		9	-829.130	1677.1	38.21	0.000
20	29.64	+	+			+	10	-828.182	1677.4	38.51	0.000
28	29.62	+	+		+	+	12	-826.444	1678.3	39.47	0.000
19	29.12		+			+	7	-840.756	1696.0	57.15	0.000
12	29.72	+	+		+		10	-837.677	1696.4	57.50	0.000
4	29.68	+	+				8	-839.946	1696.6	57.68	0.000
27	29.12		+		+	+	9	-839.047	1696.9	58.05	0.000
22	30.46	+		+		+	10	-840.246	1701.5	62.63	0.000
30	30.49	+		+	+	+	12	-838.294	1702.0	63.17	0.000
11	29.15		+		+		7	-849.819	1714.1	75.27	0.000
3	29.13		+				5	-851.946	1714.2	75.29	0.000
29	29.41			+	+	+	9	-851.724	1722.3	83.40	0.000
21	29.55			+		+	7	-854.916	1724.3	85.47	0.000
6	30.38	+		+			8	-857.291	1731.2	92.37	0.000
14	30.45	+		+	+		10	-856.468	1734.0	95.08	0.000
5	29.49			+			5	-869.465	1749.2	110.32	0.000
13	29.45			+	+		7	-867.805	1750.1	111.25	0.000
26	30.69	+			+	+	10	-865.075	1751.2	112.29	0.000
18	30.50	+				+	8	-870.047	1756.8	117.88	0.000
10	30.58	+			+		8	-873.011	1762.7	123.81	0.000
25	29.69				+	+	7	-877.369	1769.3	130.37	0.000
17	29.56					+	5	-882.285	1774.8	135.97	0.000
9	29.59				+		5	-883.557	1777.4	138.51	0.000

Models ranked by AICc(x)

Table S7.1. Output of the 'dredge' function fitting the full model and models with terms removed.

Using the function 'dredge' in the package 'MuMIn', the model with all terms included had the lowest AICc but the increase in AICc by the removal of the Month term (delta 0.55) indicated that inclusion of this term contributed little to model improvement. Thereafter, removal of other terms substantially increased AICc (Table S7.1).

```
Family: gaussian
Link function: identity

Formula:
Cover ~ s(Decimal_date, k = 25) + s(SLA, k = 25) + fTransect +
    s(month) + s(DHW, k = 15)

Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    29.848      1.098  27.174   <2e-16 ***
fTransectTR14  -2.000      1.493  -1.340   0.1818
fTransectTR15   2.903      1.566   1.854   0.0652 .
fTransectTR9   -3.624      1.467  -2.471   0.0143 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                edf Ref.df      F  p-value
s(Decimal_date) 15.353 15.353 18.346  < 2e-16 ***
s(SLA)           2.272  2.272 11.127 1.27e-05 ***
s(month)         1.000  1.000  4.565   0.0338 *
s(DHW)           1.000  1.000 53.472  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.689
  Scale est. = 61.916    n = 227
```

Phi estimate = 0.16557
fTransect p = 0.000197

Table S7.2 Model output

All terms were significant (Table S7.2) with the full model explaining 68.9% of the variability in the data and individual terms contributing in the following proportions:

Time	44.5%
SLA	2.2%
DHW	5.4%
Month	5.4%

Partial plots of the model terms.

Inspection of the partial plots for each term (Fig S7.1) and the model output (Table S7.2) indicated that the variation with Month was linear (edf=1) and involved a slight decrease in coral cover where data was collected during the later months in the year. Given the timing of *p*IODs, which occur in the boreal autumn, this relationship is not wholly unexpected since we expect a *p*IOD to induce a drop in sea level. Although the contribution of this term to the model was marginal, it was retained in the final model.

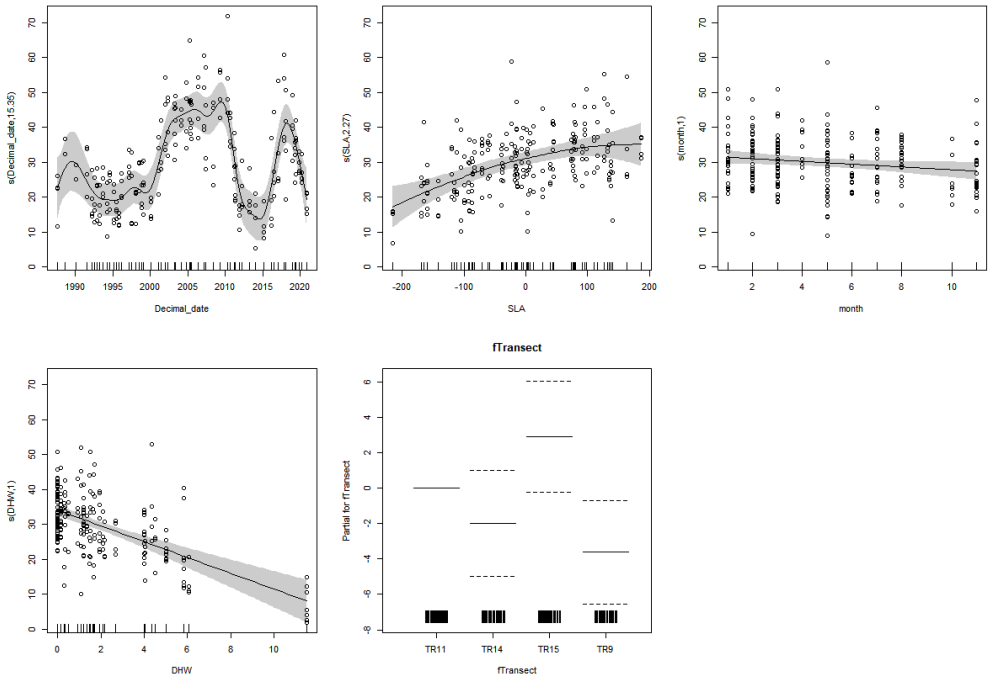

Fig S7.1. Partial effect plots for the GAMM model. (1) non-linear relationship between time and coral cover (2) curvilinear relationship between SLA and cover (3) linear relationship between month and cover (4) linear relationship between DHW and cover, and (5) Transect as a factor. Residuals plotted as open circles. Partial effects include model intercept. Shaded areas/ dotted lines are 95% CI.

<u>Model Checking</u>

Models were checked using the 'mgcv' function 'gam.check' for convergence, and the appropriate number of basis functions. The 'concurvity' function was used to check for correlation between variables. Model validation also included the following steps:

```
'gamm' based fit - care required with interpretation.
Checks based on working residuals may be misleading.
Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                  k'   edf k-index p-value
s(Decimal_date) 24.00 15.35    1.07    0.76
s(SLA)          24.00  2.27    1.03    0.64
s(month)         9.00  1.00    0.93    0.12
s(DHW)          14.00  1.00    0.84  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Although the DHW term throws up a significant p value it is clear that with a k of 14 and an edf of 1 there is not a problem.

Graphical Plotting

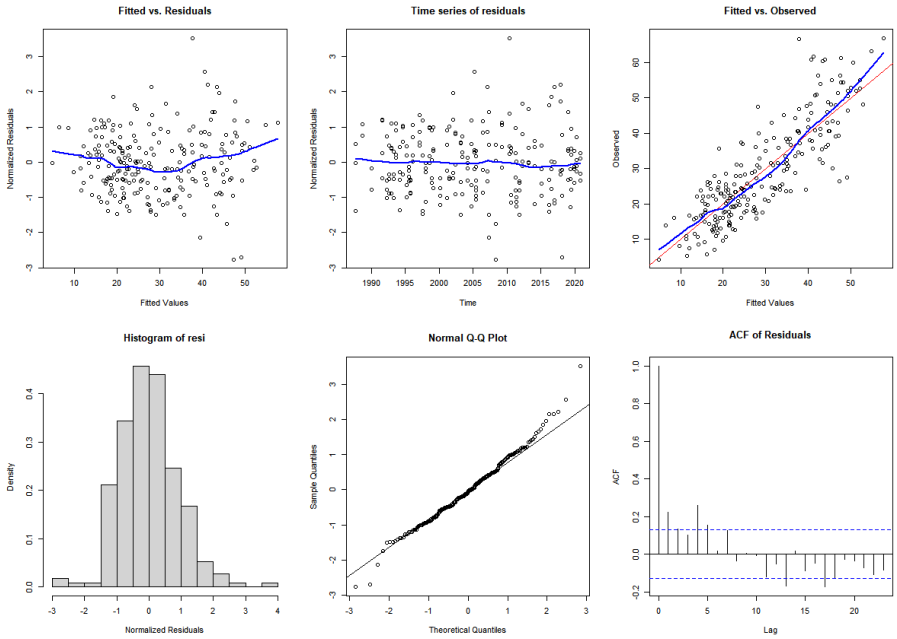

Fig S7.2. Diagnostic plots for the final model.

Residuals were normal, homoscedastic and the AR(1) term removed most of the autocorrelation (Fig S7.2).

Plots of the model residuals against each covariate (Fig S7.3) indicated that there were no patterns remaining in the residuals.

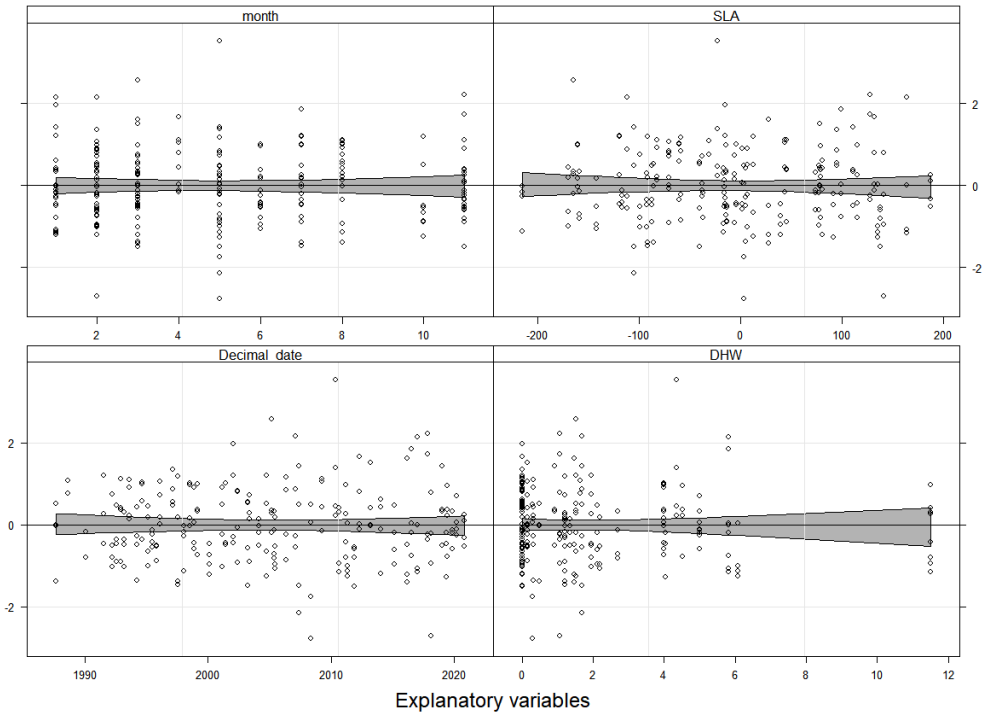

Fig S7.3. Plots of residuals versus covariates. Smoother and 95% point-wise confidence intervals (grey shade). The confidence intervals contain 0 for all covariate values indicating no significant patterns.

Finally, conditional boxplots of the residuals by Transect and month (Fig S7.4) confirmed no significant residual effect.

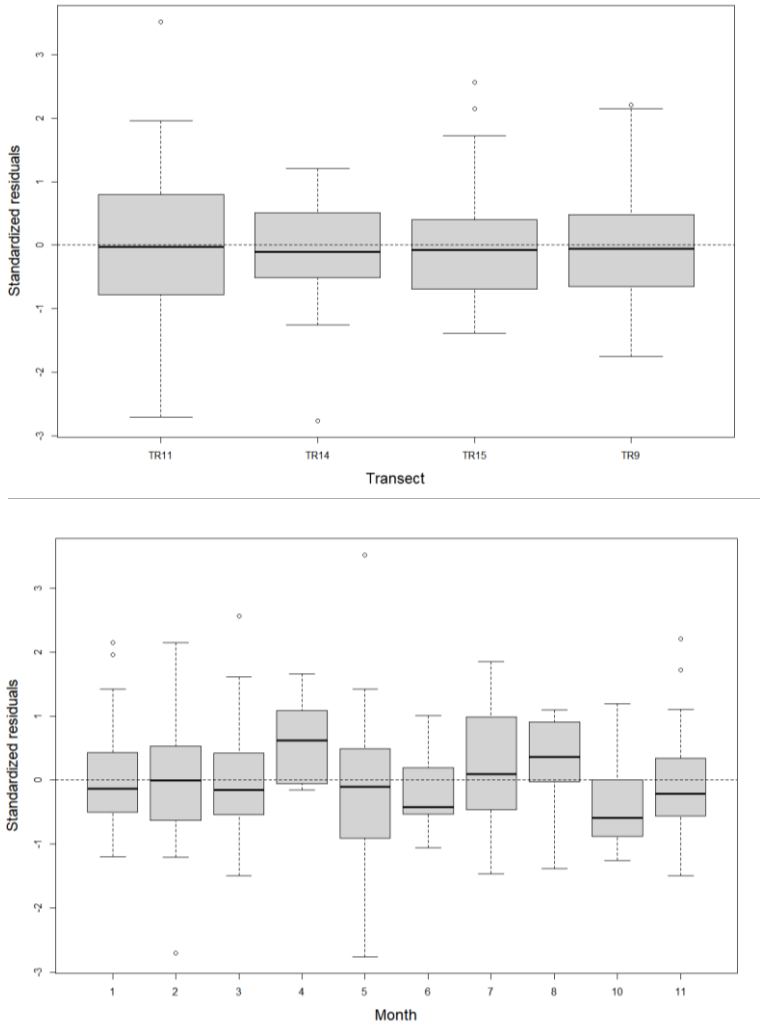

Fig S7.4. Conditional boxplots of residuals grouped by month.

Visualising the Model Output

As well as inspecting model output in the form of partial plots (Fig S7.3 above), the overall model and individual terms were extracted and plotted using the 'predict' function in 'mgcv'.

To represent the output from the model the effect of each term was examined separately as well as the overall model. Since the model is additive, the overall model is a summation of each term (time, SLA, DHW, month) together with the intercept.

The overall model fit for all terms is shown in Fig S7.5. Separate fits for each of the terms are shown in Figs S7.6 – S7.12.

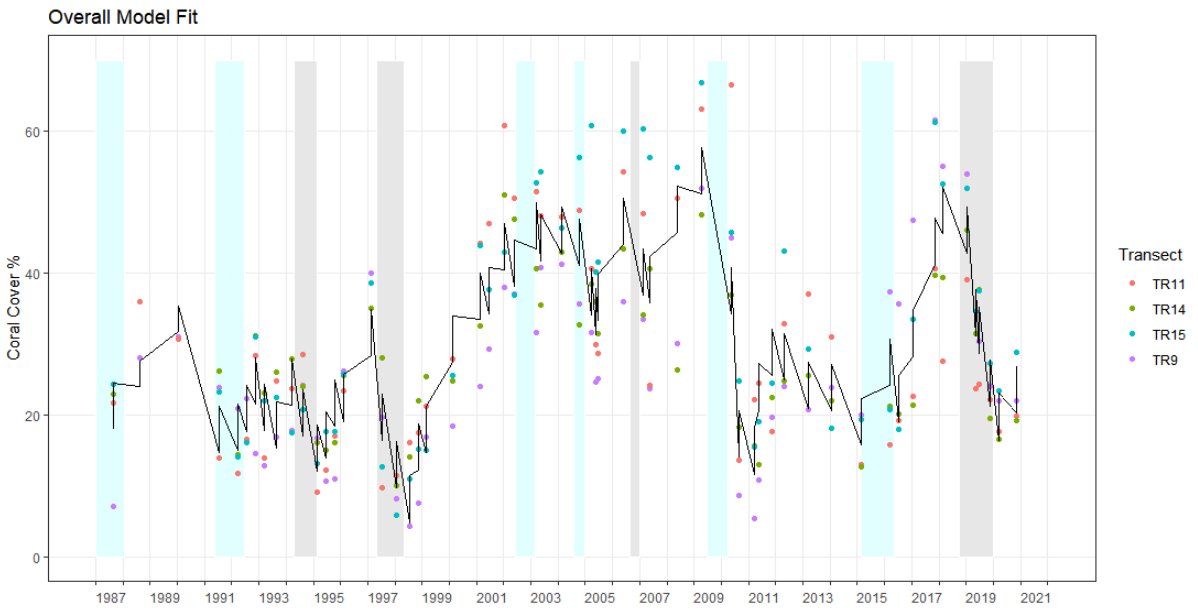

Fig S7.5. Monthly mean coral cover by transect (dots) with overall model fit (black line). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Niño (Nino3.4 > 0.5 °C); cyan shading El Niño only.

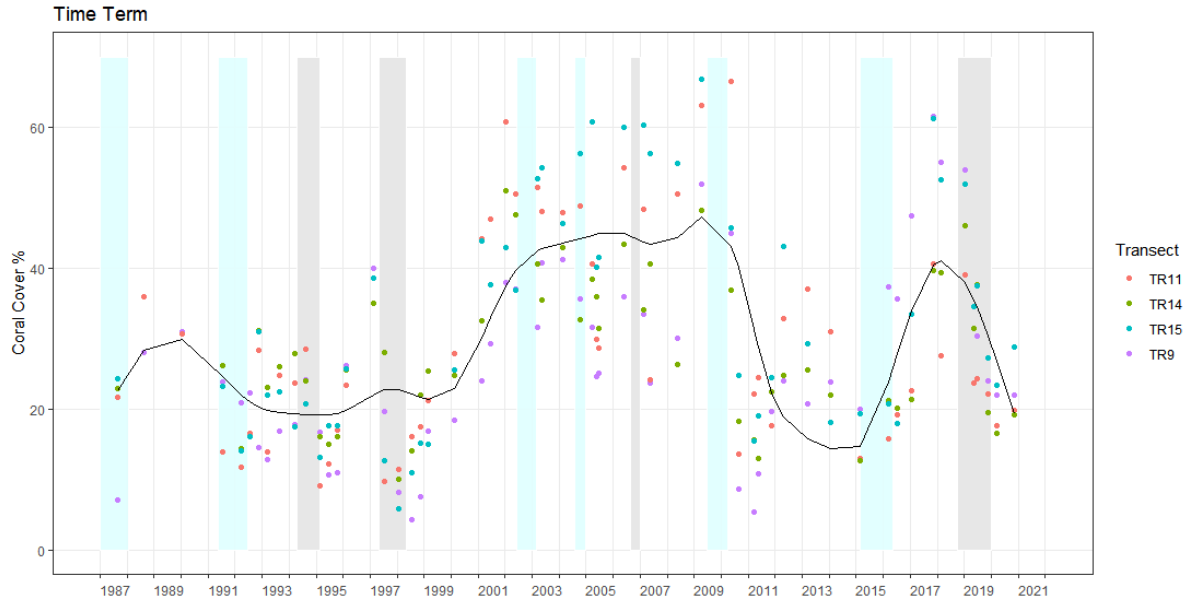

Fig S7.6. Monthly mean coral cover by transect (dots) with fit of the term for time (black line) and 95% confidence intervals (dotted lines). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Niño (Nino3.4 > 0.5 °C); cyan shading El Niño only.

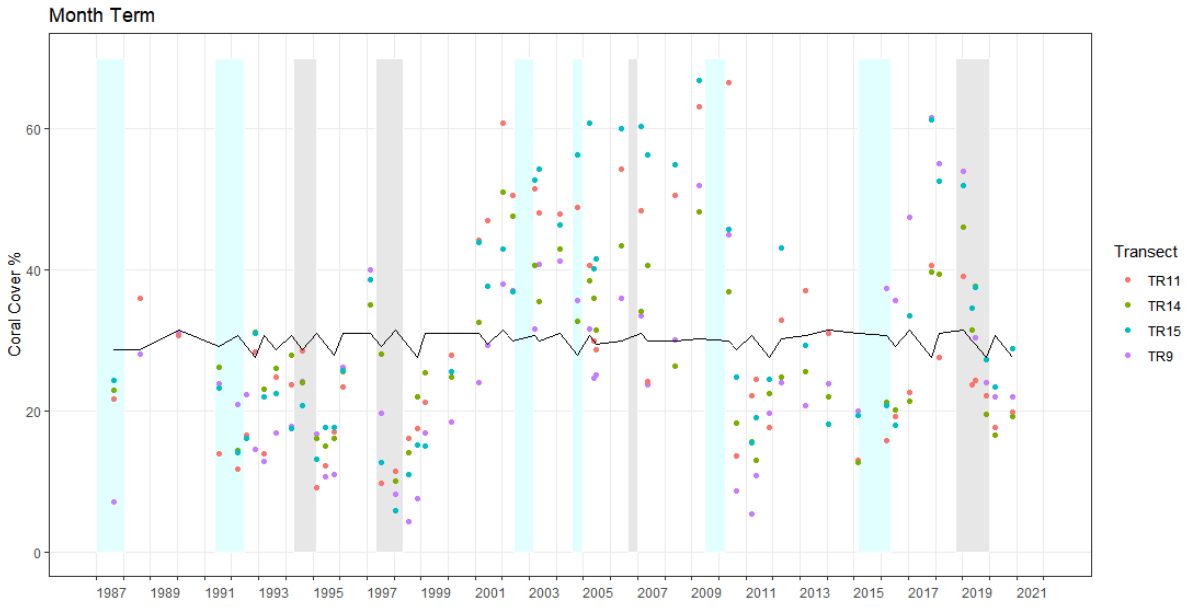

Fig S7.7. Monthly mean coral cover by transect (dots) with fit for month term (black line). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Niño (Nino3.4 > 0.5 °C); cyan shading El Niño only.

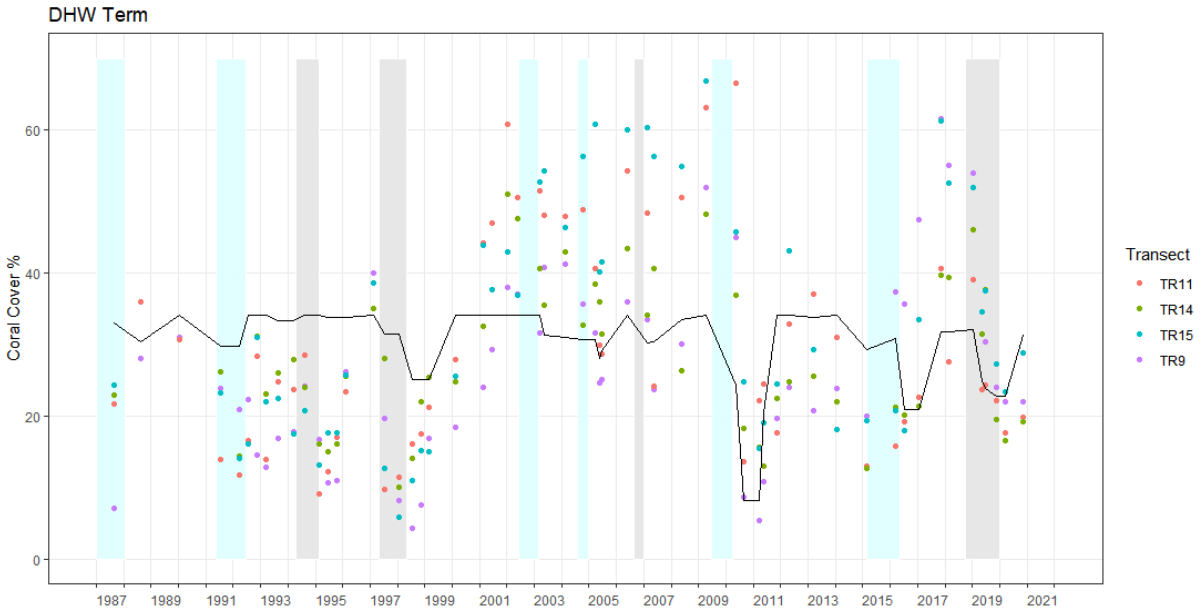

Fig S7.8. Monthly mean coral cover by transect (dots) with DHW term (black line). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Niño (Nino3.4 > 0.5 °C); cyan shading El Niño only.

The fit for DHW (Fig S6.8) shows that this term has identified the large drop in cover in 2010 when coral bleached following s sea water warming that May/June.

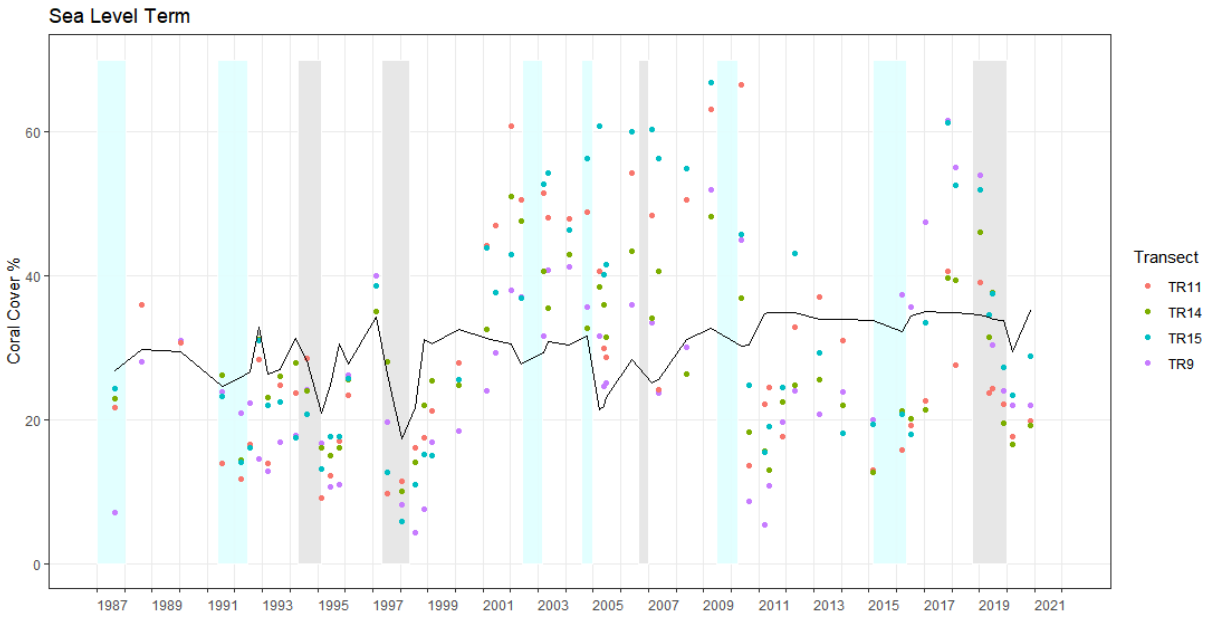

Fig S7.9. Monthly mean coral cover by transect (dots) with fit for SLA (black line). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Niño (Nino3.4 > 0.5 °C); cyan shading El Niño only.

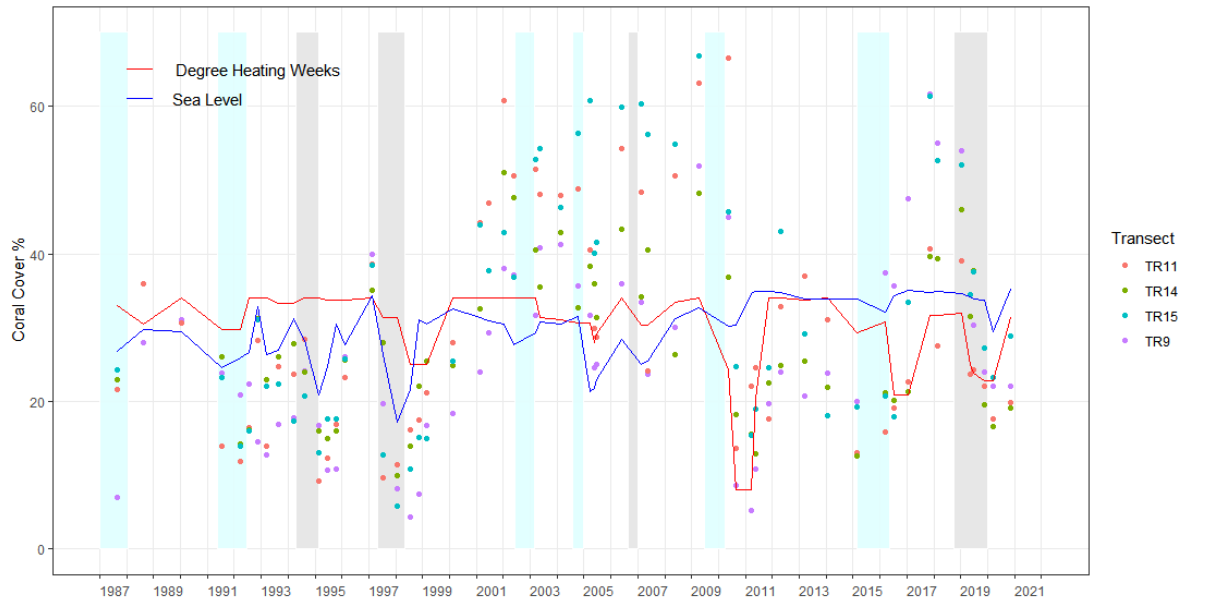

Fig S7.10. Monthly mean coral cover by transect (dots) with DHW (red line) and SLA (blue) predicted values. Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Niño (Nino3.4 > 0.5 °C); cyan shading El Niño only.

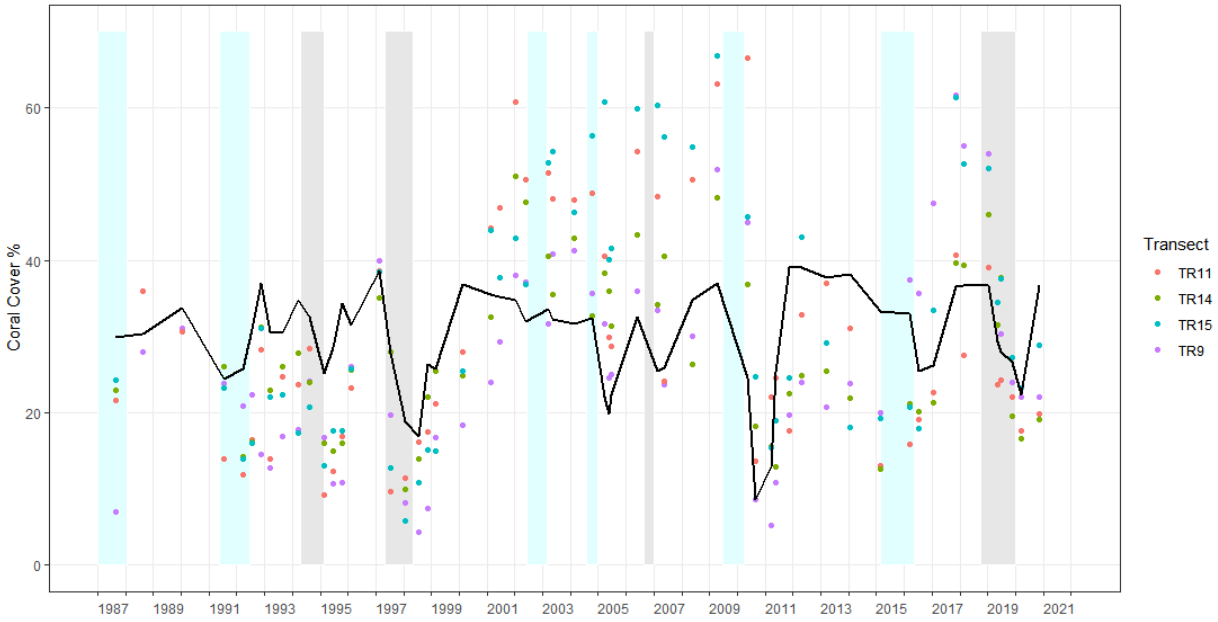

Fig S7.11. Monthly mean coral cover by transect (dots) with DHW and SLA terms combined (black line). Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Niño (Nino3.4 > 0.5 °C); cyan shading El Niño only.

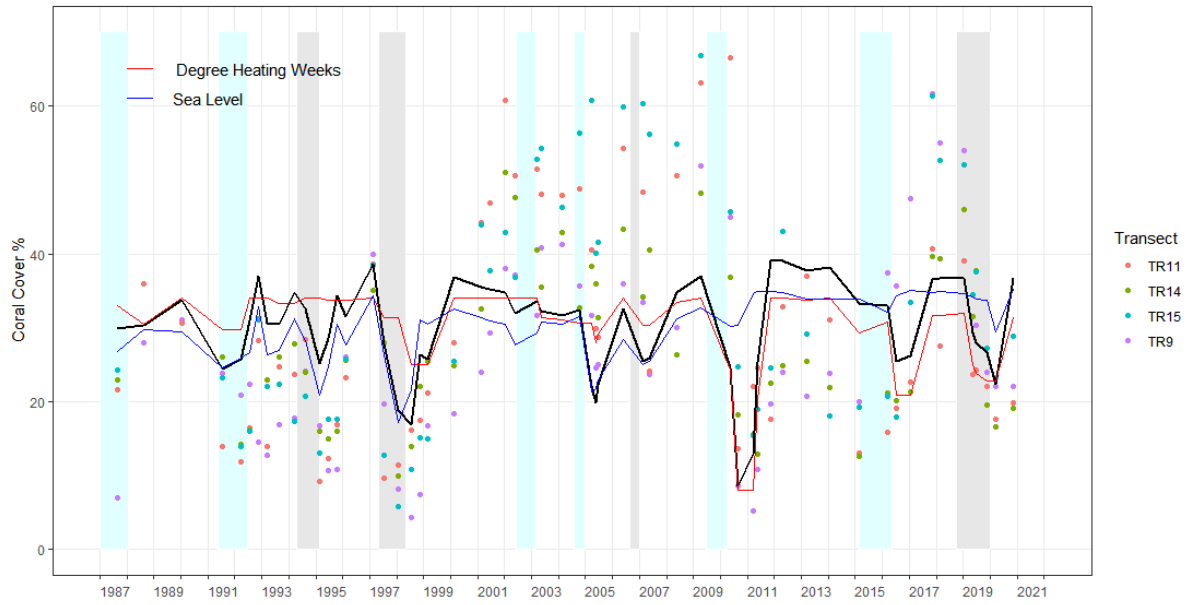

Fig S7.12. Monthly mean coral cover by transect (dots) with DHW and SLA terms combined (black line) and shown separately for comparison. Grey shaded time periods represent timing of combined *p*IOD (DMI >0.56 °C) and El Niño (Nino3.4 > 0.5 °C); cyan shading El Niño only.

References

Bamston AG, Chelliah M, Goldenberg SB (1997) Documentation of a highly ENSO-related sst region in the equatorial pacific: Research note. Atmosphere-Ocean 35:367-383

Barton K (2020) MuMIn: Multi-Model Inference. R package version 1.43.17. https://CRAN.R-project.org/package=MuMIn

Brown BE, Phongsuwan N (2012) Delayed mortality in bleached massive corals on intertidal reef flats around Phuket, Andaman Sea, Thailand. Phuket Marine Biological Centre Research Bulletin 71:43-48

Brown BE, Clarke KR, Warwick RM (2002) Serial patterns of biodiversity change in corals across shallow reef flats in Ko Phuket, Thailand, due to the effects of local (sedimentation) and regional (climatic) perturbations. Mar Biol 141:21-30

Brown BE, Dunne RP, Phongsuwan N, Somerfield PJ (2011) Increased sea level promotes coral cover on shallow reef flats in the Andaman Sea, eastern Indian Ocean. Coral Reefs 30:867-878

Ieno EN, Zuur AF (2015) A Beginner's Guide to Data Exploration and Visualisation with R. Highland Statistics, Newburgh, Scotland

Liu G, Heron SF, Eakin CM, Muller-Karger FE, Vega-Rodriguez M, Guild LS, De La Cour JL, Geiger EF, Skirving WJ, Burgess TFR, Strong AE, Harris A, Maturi E, Ignatov A, Sapper J, Li J, Lynds S (2014) Reef-Scale Thermal Stress Monitoring of Coral Ecosystems: New 5-km Global Products from NOAA Coral Reef Watch. Remote Sensing 6:11579-11606

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Saji NH, Goswami BN, Vinayachandran PN, Yamagata T (1999) A dipole mode in the tropical Indian Ocean. Nature, London 401:360-363

Wood SN (2017) Generalized Additive Models - An Introduction with R, Second Edition. Chapman and Hall/CRC

Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1:3-14

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) Mixed Effects Models and Extensions in Ecology with R. Springer, New York