Data Modelling of Marine Paleoclimate Observations for Reconstructing El Niño

Sam Hommel

Advisor: Dr. Mike Evans

GEOL 394

Abstract

To address changes in El Niño Southern Oscillation behavior in response to anthropogenic forcing, simple linear anomaly data models linking time series paleoclimate observation data sets to gridded analyzed historical climate data were developed, facilitating data assimilation. Four series of linear regressions were performed: coral $\delta^{18}O$ anomaly on sea surface temperature (T) anomaly, coral Sr/Ca ratio anomaly on T anomaly, precipitation (P) anomaly on T anomaly, and, through stepwise regression, the residual of coral $\delta^{18}O$ on T on the residual of P on T. Of 71 coral $\delta^{18}O$ time series data sets, 52 yielded significant linear models when regressed with T. Of 32 Sr/Ca sets, 6 yielded significant models when regressed with T. Of 52 P sets, 28 yielded significant models when regressed with T. Of 28 e_1 and e_3 sets, 11 yielded significant models. The median (statistically significant at $p < 0.05$) data models (95% confidence interval) of the general form:

$$Y = (h)*(X) + e$$

where Y is the dependent variable, X is the independent variable, h is the linear coefficient linking the two, and e is the residual, are as follows:

$$\delta^{18}O = (-0.22\ [-0.32, -0.14])*(T) + e_1$$

$$Sr/Ca = (-0.06\ [-0.09, -0.02])*(T) + e_2$$

$$P = (56.37\ [24.33, 108.32])*(T) + e_3$$

$$e_1 = (-0.0011\ [-0.0019, -0.00044])*(e_3) + e_4$$

The median regression coefficients for the $\delta^{18}O$ on T and Sr/Ca on T models are consistent within uncertainty with regression coefficients independently derived from manipulative and space-for-time experiments. Each significant data model connecting $\delta^{18}O$ and Sr/Ca to T was

employed in and validated through a data assimilation experiment using the Zebiak-Cane (1987) model.

Table of Contents

Introduction

Through global, interconnected climate patterns, the natural variability of the El Niño Southern Oscillation (ENSO) drives changes in temperature and precipitation all over the world (Collins et al. 2010). ENSO has become, through this wide climatic influence, and its eminent predictability, a key part of the initialization of many global climate models (GCM) (Chen et al. 2008). As climatic analyses continue to point to anthropogenic climate change as a significant and relatively novel source of climate forcing (Charlson et al. 1992), it is vital to improve the reconstructive skill of ENSO models to adequately address changes in ENSO as a result of these new forcings. Previous projects (Chen et al. 2004, Steiger et al. 2018) have demonstrated the power of data assimilation (DA), a comparison between model results and corresponding observational data, when applied to models of ENSO, but modern climate data is inherently influenced by anthropogenic forcings (Charlson et al. 1992). To achieve an independent comparison of ENSO behavior before and after the introduction of anthropogenic climate influence, DA between paleoclimate observations and model results was performed, facilitated by data modelling.

ENSO is a powerful, internally driven climatic phenomenon whose primary indicators can be seen in the tropical Pacific: anomalously low sea surface temperature (T) and precipitation (P) in the eastern tropical Pacific for a “La Niña” state, and anomalously high T and P in the same region for an “El Niño” state. These states develop, to first order, through the Bjerknes feedback: during a La Niña state, the eastern Pacific sea surface is colder than the western Pacific sea surface, inducing an easterly trade wind, which drives an upwelling of cold water from the deep ocean in the eastern Pacific, intensifying the La Niña state (Collins et al.

2010). The weakening of the easterly trade wind during an El Niño state leads to its own positive feedback: when surface winds are weaker, they drive a weaker easterly sea surface current, meaning that warm surface waters stagnate in the eastern Pacific, feeding back into the El Niño ENSO state. Transitions between the states are often driven by slower and more intricate processes, such as the gradual discharge of tropical Pacific surface water reservoirs poleward, and small-scale instabilities within the system (Collins et al. 2010).

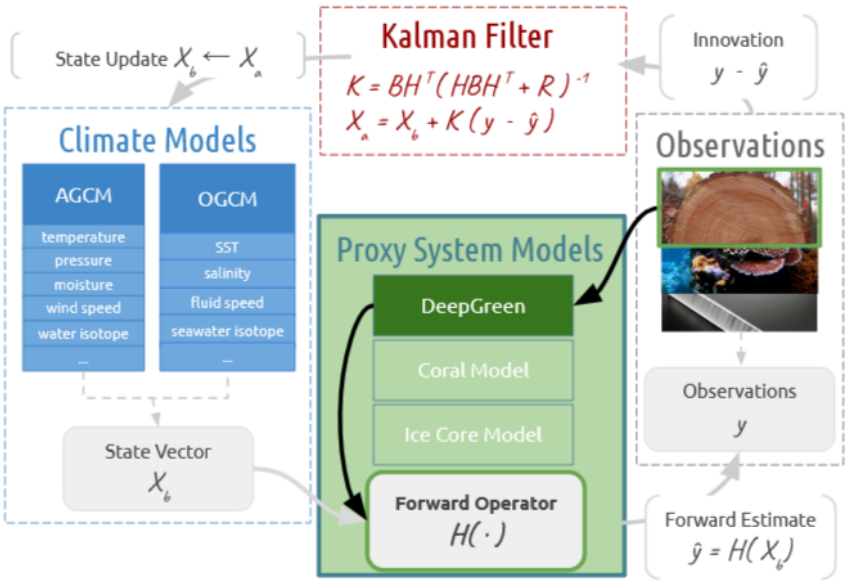

Figure (1): A schematic of a climate model DA framework adapted from Hakim et al. (2016). Note the flow from model state vector (variable) on the left, through the forward operator and estimate, towards observations. Note also the necessity of observations to develop the forward operator, as indicated by the black arrows.

DA, at the coarsest level, seeks to update model results based on observational data to achieve an improved estimate relative to one based on either data or simulations separately. A climate model, a mathematical representation of how a real-world process changes with time, is defined by multiple state variables. Climate models produce values for these state variables at certain times which correspond with observable data of the same variable at the same time. DA systems reduce the uncertainty of model predictions by combining model results and observations using their respective uncertainties.

As shown in Figure 1, the DA process relies on proxy system modelling (PSM), and more specifically, a forward operator and estimate. These tools can be used find a solution that minimizes the misfit between the climate state, as mapped the forward operator ("H" in the diagram), and the observations. Figure 2 is a rough schematic of a PSM framework. PSM frameworks rely on a series of models, describing the way a sensor interprets its environment, the way that sensor archives environmental information, and the way the archive is accessed through sampling. These three steps can be approximated by implementing a linear data model. For the purposes of DA, data models are commonly framed as forward models, meaning they predict changes in a dependent variable (often corresponding to observations made on the archive of a PSM) based on changes in an independent variable (often the environmental component of a PSM).

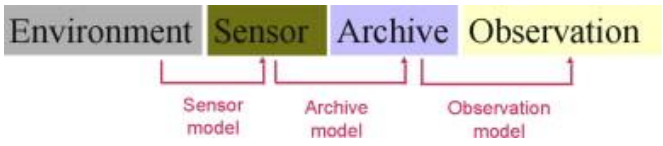

Figure (2): Outline of a PSM framework. (Evans et al. 2013)

The goal of this project was to develop simple but validated data models linking paleoclimate observations as the dependent variable to climate vectors as the independent variable, to facilitate the DA of marine paleoclimate observations into the Zebiak-Cane (ZC) model of ENSO. By reconstructing ENSO with paleoclimate observations, its behavior prior to the introduction of anthropogenic forcing can be estimated, allowing for an independent comparison with modern ENSO behavior.

The model of focus for this study was the ZC coupled ocean-atmosphere anomaly model (Zebiak and Cane 1987). Specifically, this study focuses on the 5th Lamont-Doherty Earth Observation Network (LDEO5) version of the ZC model. The ZC model is of intermediate complexity, but still captures much of the intricacy of the ENSO system through its simulation of

ocean-atmosphere interactions. The ZC model domain is restricted to the tropical Pacific, the region in which ENSO events are most clearly indicated (Collins et al. 2010). The ZC model has demonstrated high forecast skill through DA, able to accurately predict ENSO behavior at leadtimes of up to 2 years (Chen et al. 2004). It simulates a variety of variables, including T anomaly, and atmospheric heating (AH) anomaly, which are both indicators of ENSO behavior, and are thus the independent variables this study used to develop data models.

With independent variables of T and AH, dependent variables in the form of paleoclimate observations were also necessary to develop simple linear data models for DA. An important archive for paleo-environmental data corresponding to these variables in the tropical Pacific is aragonite in near-surface-dwelling corals (Steiger et al. 2018). These corals act as sensors in a PSM framework, archiving changes in environmental variables as chemical signatures in their carbonate skeletons (Evans et al. 2013). Two such chemical signatures are $\delta^{18}O$ and Sr/Ca ratio.

Coral $\delta^{18}O$, the ratio of ^{18}O to ^{16}O in coral aragonite, changes as a function of calcification temperature (T_c) (McCrea 1950, Epstein et al. 1953), which can be equated, with low uncertainty, to environmental temperature (T) (O'Neil et al. 1969). Previous studies constrained the relationship between coral $\delta^{18}O$ and T using simple linear models similar to those derived as part of this project, with linear coefficients of: -0.183 permil/°C (McCrea 1950), and -0.23+/-0.02 (2σ) permil/°C (Epstein et al. 1953). These coefficients describe a relationship such that, as temperature of the solution increases, the $\delta^{18}O$ of aragonite decreases.

Coral $\delta^{18}O$ can also be related to the $\delta^{18}O$ of seawater surrounding the coral (Grossman and Ku 1986, Thompson et al. 2011). The $\delta^{18}O$ of seawater is affected by a wide variety of variables, including T and P (Thompson et al. 2011). By reducing collinearity from T, coral $\delta^{18}O$ can be used as a paleoclimate observation corresponding to P. P is not explicitly simulated in the

ZC model, so it must be related to AH. AH as modelled in ZC can be divided into two components, one dependent on anomalous surface temperature, and one dependent on latent heat and moisture convergence (Zebiak and Cane 1987). Based on latent heat gain associated with increased P (as more clouds condense, more latent heat is generated), P and AH can be related (Chiang et al. 2001), finally connecting the paleoclimate observation of coral $\delta^{18}O$ to the state variable AH.

Coral Sr/Ca ratio, similar to coral $\delta^{18}O$, can be linearly related to T_c (Kinsman and Holland 1969, Gagan et al. 2000). These previous studies produced linear coefficients describing the linear relationship between the observation and state variable: -0.09+/-0.03(2σ) mmol/mol/°C (Kinsman and Holland 1969), and -0.062 +/- 0.014 (2σ) mmol/mol/°C (Gagan et al. 2000). As with coral $\delta^{18}O$, coral Sr/Ca ratio tends to decrease as T increases.

Simple linear data models, which approximate the complex relationships found in PSM, have demonstrated their power as predictors of climate behavior in the ENSO region (Epstein et al. 1953, Grossman and Ku 1986, Gagan et al. 2000), and are necessary for a crucial step in data assimilation of ENSO models (Figure 1). Many of these models follow, with variations, this general form:

$$D = h_1 * (V_1) + e_1 \quad (1)$$

where D represents paleoclimate observations, V_1 represents climate variables, h_1 describes the linear relationship between the two, and e_1 signifies the residual (the component of D that is not dependent on V_1).

This simple model can be expanded, made more complex, in order to capture more variation in the modelled process. This can be accomplished through stepwise linear regression,

which relies on the residuals of a series of univariate least squares regressions. Suppose there is a second climate variable, V_2, that is linearly related to V_1:

$$V_2 = h_2 * (V_1) + e_2 \quad (2)$$

where V_2 is the dependent variable, V_1 is the independent variable, h_2 captures the linear relationship between the two, and e_2 describes the component of V_2 which does not covary with V_1. This final regression model:

$$e_1 = h_3 * (e_2) + e_3 \quad (3)$$

constrains the relationship between the residual of the D/V_1 regression versus the V_2/V_1 regression. Equation 3 defines the linear relationship between D and V_2, accounting for confounding influence from V_1.

In this study, I applied the general data model framework laid out in Equations 1-3 using paleoproxy and climate data to deepen understandings of ENSO behavior and the chemical relationships linking proxies to environmental influences, as well as facilitate ZC model DA. I used the general model shown in Equation 1 to analyze the relationship between coral $\delta^{18}O$ and T, and coral Sr/Ca ratio and T. I used the stepwise framework described by Equations 2 and 3 to constrain the reliance of coral $\delta^{18}O$ on P, accounting for the influence of T.

Hypothesis

I hypothesized that gridded time series data sets of tropical Pacific paleoclimate proxy data, specifically coral $\delta^{18}O$ and Sr/Ca ratio, when regressed with Zebiak-Cane model associated gridded historical climate observations of sea surface temperature and precipitation, will yield significant, linear data models with non-zero slope coefficients, allowing for data assimilation.

Statistical hypotheses were as follows:

$$\delta^{18}O \text{ (permil)} = (h_1)*(T[°C]) + e_1,$$

$$\text{Null: } h_1 = 0$$

$$P_{crit} = 0.05$$

$$\text{Sr/Ca (mmol/mol)} = (h_2)*(T[°C]) + e_2$$

$$\text{Null: } h_2 = 0$$

$$P_{crit} = 0.05$$

$$\text{P (mm/month)} = (h_3)*(T[°C]) + e_3$$

$$\text{Null: } h_3 = 0$$

$$P_{crit} = 0.05$$

$$e_1\text{(permil)} = (h_4)*(e_3[\text{mm/month}]) + e_4$$

$$\text{Null: } h_4 = 0$$

$$P_{crit} = 0.05$$

Data/Methods

I sourced paleoclimate observation data from a compilation by Steiger et al. (2018), which they used to develop their Paleo Hydrodynamics Data Assimilation (PHYDA) product. The PHYDA database is comprised of 2978 discrete proxy data time series, from years 1-2012. Many series were derived from the PAGES2k database (Emile-Geay et al. 2017), a compilation drawing on coordinated paleoclimatology community involvement. After restricting locations to the ZC model basin (28.75°S to 28.75°N, 124°E to 80°W) and to marine observations, I was left with 71 coral $\delta^{18}O$ data sets, and 32 coral Sr/Ca ratio data sets.

I obtained historical gridded analyzed sea surface temperature (T) data based on the locations of the paleoclimatic observations from Kaplan et al. (1998-2000). This database used

simplified descriptions of T covariance in space and time to estimate T at locations for which there were no direct observations, at a monthly resolution from 1856-1991.

For use in stepwise regressions, I gathered precipitation data at the coral sites from a compilation by Baker et al. (1995). The compilation includes both temperature and precipitation data, and is based on quality controlled data drawn from the Global Historical Climate Network (GHCN). Baker et al. (1995) refined patchy and uncertain raw data through objective data analysis: a two-step procedure that (1) flagged and filtered out significant outliers, and (2) utilized spatial interpolation to estimate temperature and precipitation values for periods lacking data, within the period 1850-1993.

I performed each regression using MATLAB's "regress" function. Considering the model of focus for this study (ZC) is an anomaly model, I formed δ^{18}O, Sr/Ca, T, and P anomalies by subtracting the mean for the interval: 1901-2000. As such, the "y-intercept" common to many simple linear relationships is irrelevant for this study, so I focus only on the regression coefficient (h_n). Based on assumptions of stationarity, a concept explored further in the discussion section, the results for the climate of this period can be applied to paleoclimates with acceptable uncertainty. To facilitate significance testing, critical f-values ranging from 3.89 to 3.94 based on degrees of freedom ranging from 51-99 were selected for each regression.

Results:

The significant results of the regression of coral δ^{18}O on local T are displayed in Figure 3; all results are in Table A1 (Appendix). Of 71 coral δ^{18}O data sets drawn from PHYDA, 52 yielded significant data models when regressed with T. The median (95% confidence interval) of the significant data models was:

$$\delta^{18}O = (-0.22\ [-0.32, -0.14])*(T) + e_1$$

(4)

Results of Sr/Ca ratio on local T, are shown in Figure 4 and Table A2. Of 32 total Sr/Ca data series regressed against local T, 6 displayed a significant coefficient. The median (95% confidence interval) of the significant data models was:

$$Sr/Ca = (-0.06\ [-0.09, -0.02])*(T) + e_2$$

(5)

The significant results of the regression of P on local T are displayed in Figure 5, all coefficients can be found in Table A3. Of the 100 total data sets, 28 produced significant data models, which are displayed in Figure 5. The median (95% confidence interval) of the significant data models was:

$$P = (56.37\ [24.33, 108.32])*(T) + e_3$$

(6)

The final median (95% confidence interval) data model, which captures the relationship between $\delta^{18}O$ and P while accounting for collinearity from T through stepwise least squares regression is:

$$e_1 = (-0.0011\ [-0.0019, -0.00044])*(e_3) + e_4$$

(7)

where e_1 and e_3 are the residuals of the $\delta^{18}O$/T ($\delta^{18}O_r$) and P/T (P_r) regressions, respectively. Given that the data used to derive this model were based on only the significant results of the first and third regressions, the total number of regressions performed to develop this model was 28. Of those 28 least squares relationships, 11 were significant. Significant results for this model are shown in Figure 6, all coefficients are in Table A4. Median slope coefficients at 95% confidence intervals are displayed for each regression in Table 1.

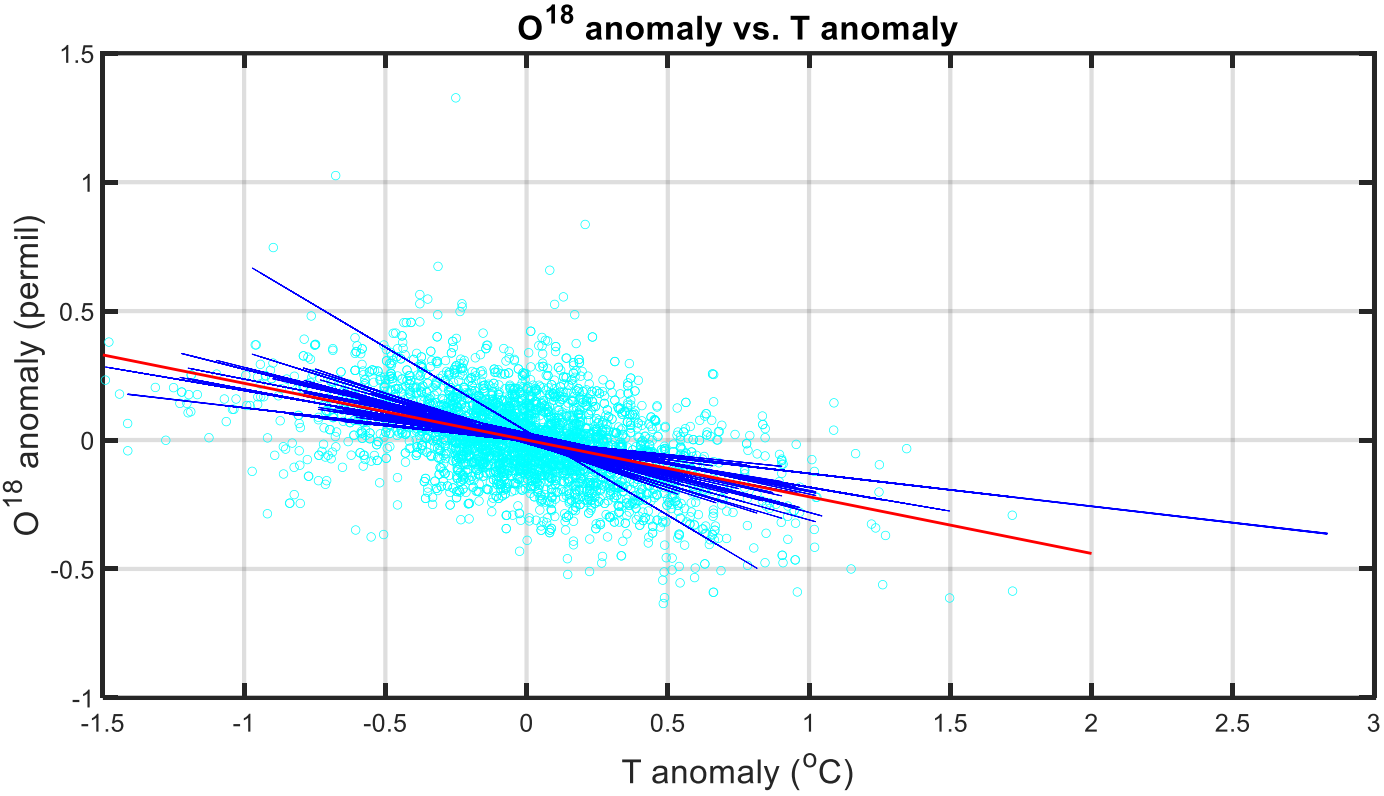

Figure (3): Results of the first data model. Cyan circles indicate $\delta^{18}O/T$ time series pairs which produced a significant regression. The blue lines are visualizations of each significant h_1. The red line represents the median h_1, using the median data model with 100 equally spaced values from T = -1.5 °C to T = 2 °C acting as the independent variable input.

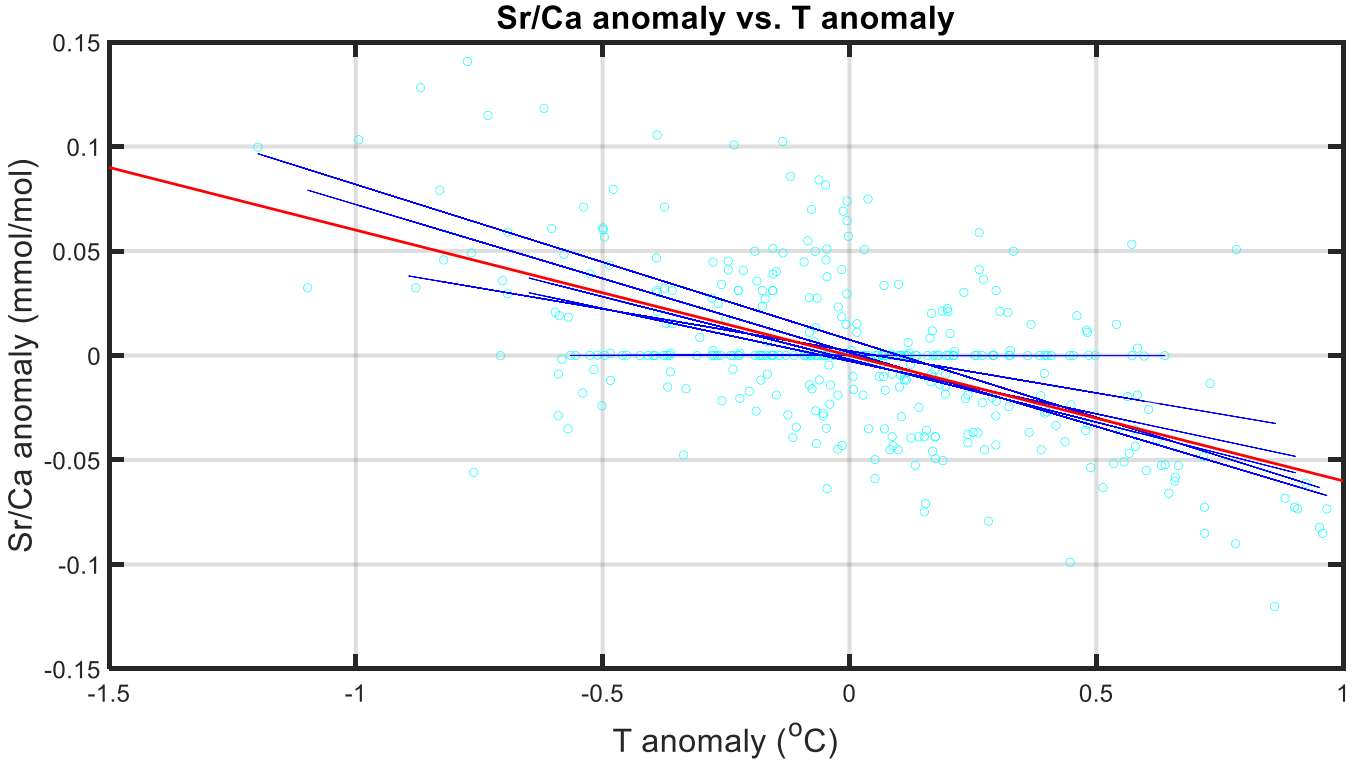

Figure (4): Results from the second data model. Symbology is the same as Figure 3, here the median slope line is based on h_2 and a T range from -1.5 °C to 1 °C.

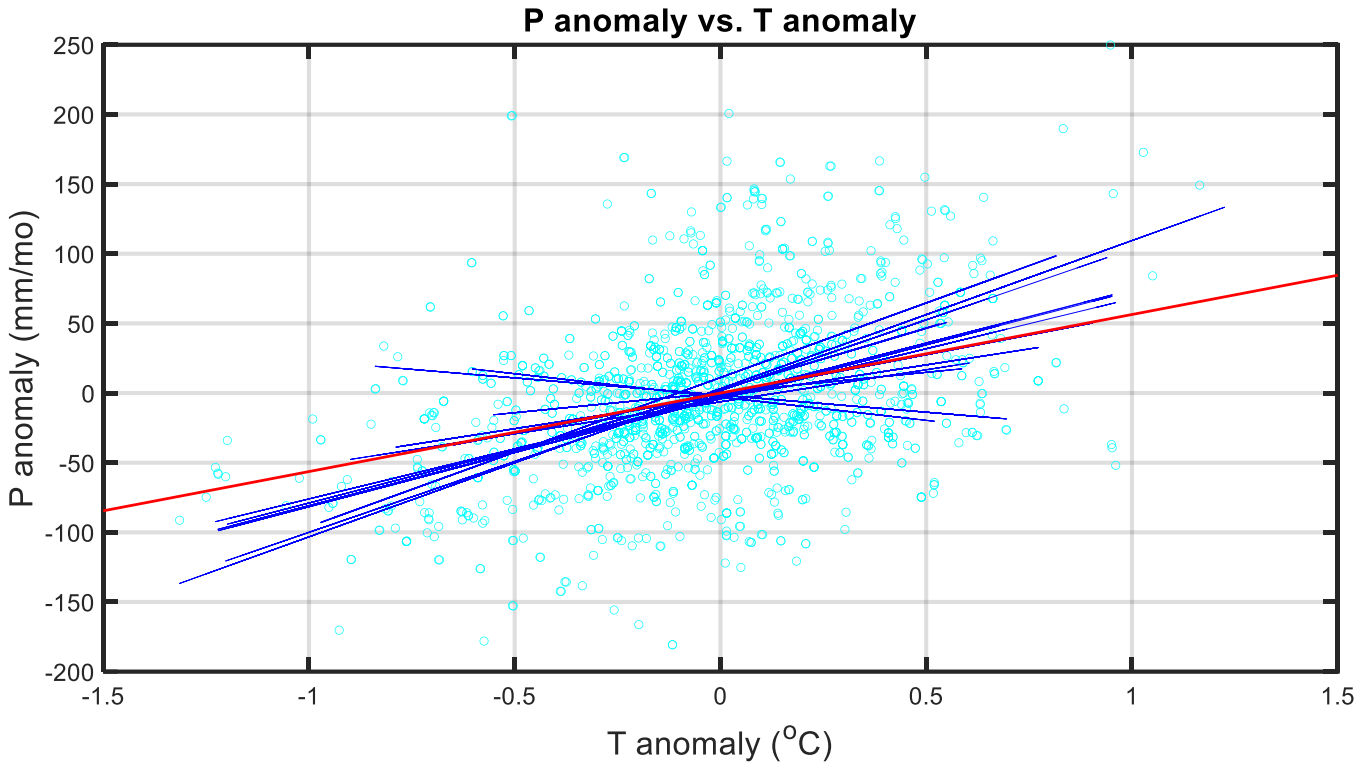

Figure (5): Results from the third data model. Symbology is the same as Figure 3. The median slope line demonstrates h_3 and is based on a T range from -1.5 °C to 1.5 °C.

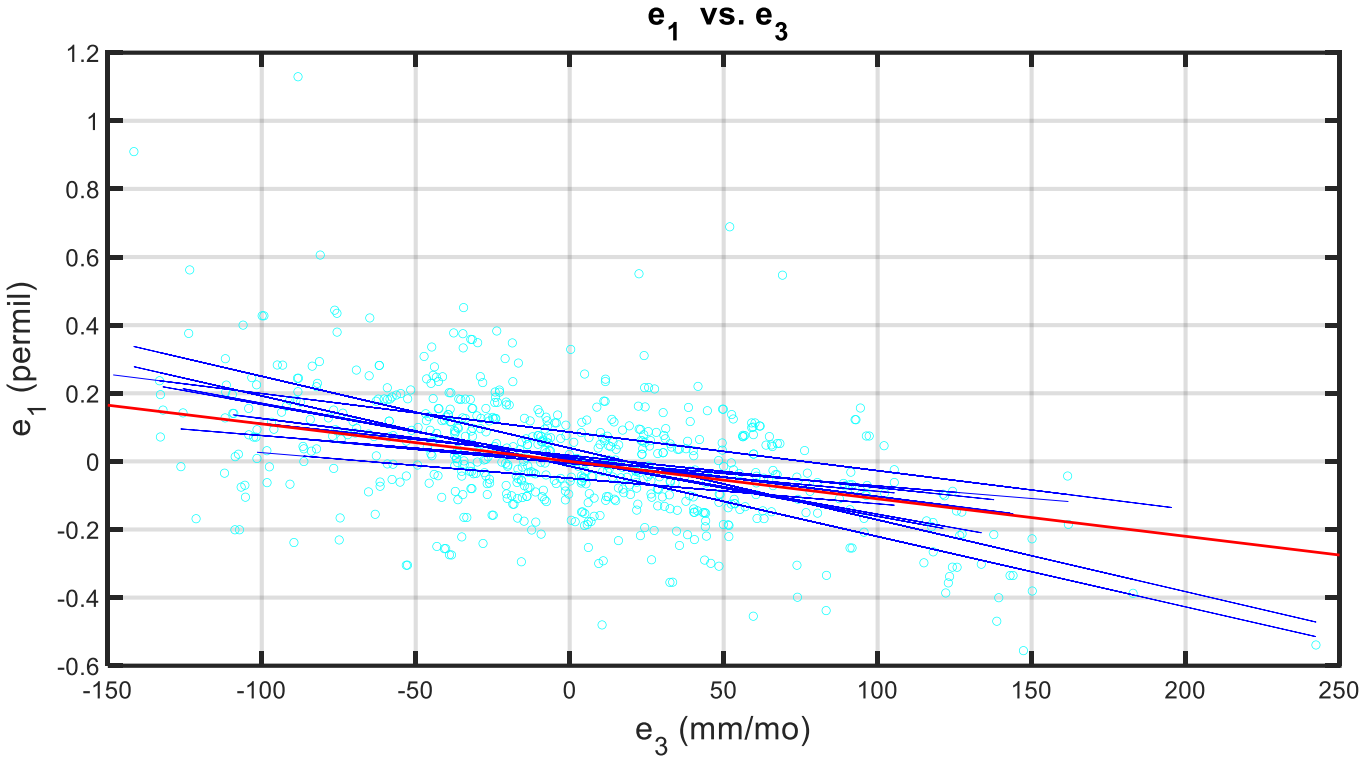

Figure (6): Results from the fourth data model. Symbology is as before. Here, the median slope line is based on h_4, and a range of P residual values from -150 mm/month to 250 mm/month

$\delta^{18}O/T$ (h_1)	Sr/Ca/T (h_2)	PPT/T (h_3)	$\delta^{18}O/PPT$ (h_4)
-0.22 (-0.32, -0.14) permil/°C	-0.06 (-0.09, -0.02) mmol/mol/°C	56.37 (24.33, 108.32) mm/month/°C	-0.0011 (-0.0019, -0.00044) permil/mm/month

Table (1): Median slope coefficients derived from each analysis. Uncertainties are reported at 95% confidence interval

Discussion:

Linear Dependence of $\delta^{18}O$ on T:

My results regarding the linear dependence of coral $\delta^{18}O$ on T (Equation 4; Figure 3; Table 1; Table A1) are consistent with prior theoretical predictions and empirical evidence for the dependence of equilibrium oxygen isotopic fractionation as a function of T_c.

Beginning with the work of McCrea (1950), geochemical studies have attempted to constrain the viability of carbonate $\delta^{18}O$ as a paleothermometer. McCrea (1950) predicted that increasing temperatures would lead to reduced oxygen isotope fractionation in inorganically precipitated carbonates based on the theoretical relationship between the two variables. Theoretically, the fractionation of oxygen isotopes in carbonates relies on the equilibrium constant of the reaction between the carbonate and surrounding water, and changes in lattice motion of the precipitated crystal, which follow changes in temperature (McCrea 1950). Through manipulation of lattice energy relations and chemical analyses of inorganic laboratory precipitated calcites and aragonites, McCrea derived a linear coefficient: -0.183 permil/°C (McCrea 1950, his Table XI), which describes the fractionation of oxygen isotopes in inorganically precipitated carbonates as a function of the temperature of the surrounding solution.

Epstein et al. (1953) performed manipulative experiments to constrain the isotopic fractionation in biogenic carbonates as a function of ambient water temperature. Their procedure involved damaging the shells of carbonate producing species, primarily abalone, and controlling the temperature at which the organisms regrew the damaged portions. They then sampled and chemically analyzed the regrown portions. Additionally, they analyzed successive growth sections of marine shells in which seasonal temperature variation was known. Epstein et al. (1953, their Table 7) reported T_c and $\delta^{18}O$ of aragonite, corrected for the $\delta^{18}O$ of the ambient seawater, over the T anomaly range: -2°C to 3°C. From their data, displayed in Figure 7, and reported in their Table 7, I estimated h from the regression of reported coral $\delta^{18}O$ (corrected for seawater $\delta^{18}O$) on T as: -0.23+/-0.02 (2σ) permil/°C. Epstein et al. asserted that from the agreement of this estimate of the biogenic $\delta^{18}O$ dependence on T_c and the results from inorganic

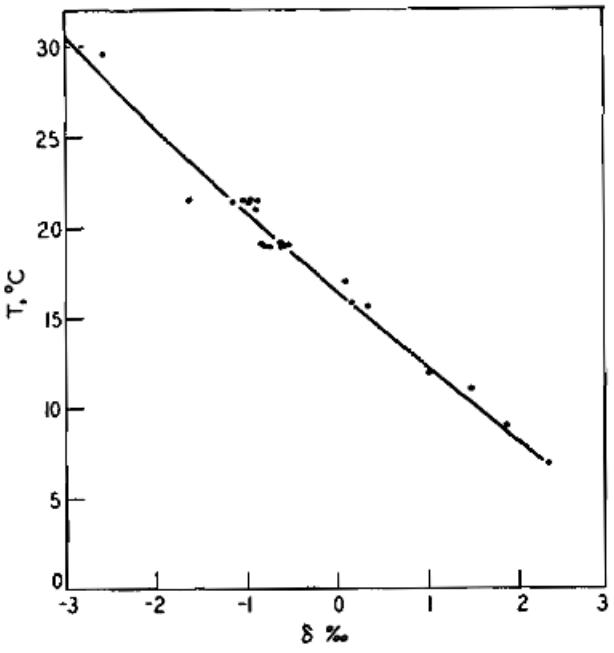

Figure (7): T plotted against $\delta^{18}O_c$, corrected for $\delta^{18}O_{sw}$. Points represent isotopic analyses, while the line is a least-squares regression. Not in anomaly units, so cannot be directly compared to figures from this study. (Epstein et al. 1953)

experiments of McCrea (1950), a valid isotopic paleotemperature could be reconstructed, independent of carbonate origin.

Further reinforcing the chemical relationship between $\delta^{18}O$ and T, Grossman and Ku (1986) performed a mensurative experiment analyzing the relationship between the isotopic fractionation of oxygen in the aragonite portions of Pacific mollusks and foraminifera as a function of water temperature, employing depth of formation as a rough temperature gradient. They, too, described a negative linear relationship between $\delta^{18}O$ and temperature in biogenic carbonate:

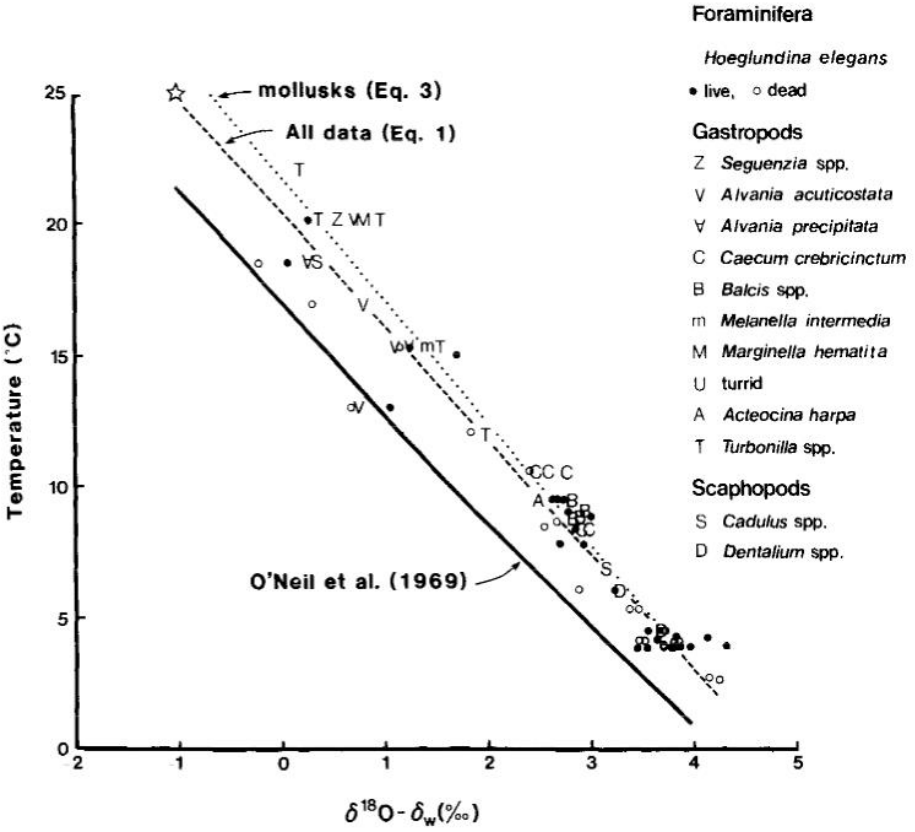

Figure (8): T vs. seawater-corrected $\delta^{18}O$ for a variety of carbonate-producing species. The least-squares line labelled "All data (Eq. 1)" corresponds to Equation 8 here. (Grossman and Ku 1986)

$$T\,(^{\circ}\mathrm{C}) = 20.6 - 4.34(\delta^{18}\mathrm{O_{ar}} - \delta_{\mathrm{w}}) \qquad (8)$$

A graphical representation of this equation can be found in Figure 8. Grossman and Ku (1986) concluded that the isotopic temperature scales displayed by the variety of species that they

analyzed agreed with previous, inorganically derived carbonate isotopic temperature scales (O'Neil et al. 1969), further implying that carbonate $\delta^{18}O$ is a valid isotopic temperature scale.

The methodology and experimental design of each of these previous works were independent from one another, and independent from the methodology and experimental design of this study. McCrea (1950) performed manipulative analyses on inorganic carbonates, Epstein et al. (1953) analyzed organic aragonite through a manipulative experiment, Grossman and Ku (1986) performed a mensurative analysis on a variety species' aragonite shells using a space-for-time temperature gradient, and this study developed isotopic temperature scales relating time series of anomaly coral $\delta^{18}O$ and T data across the Pacific, through mensurative, entirely statistical methodology. Despite independent methods and experimental design, each of these projects arrived at similar conclusions: my results (Equation 4; Figure 3; Table 1; Table A1) and those reported in McCrea (1950) and Epstein et al. (1953) agree within uncertainty. The results of this study suggest that the data model coefficients describing the relationship between $\delta^{18}O$ and T, estimated independently and using time series data, are consistent with prior results that are based on space-for-time temperature gradients, manipulative studies, and theory. Further, the results suggest the coefficients are stationary, which is important for reconstruction of T into the past.

This relationship appears to persist despite changes in the environment and manner in which the carbonate is formed, meaning that assumptions of stationarity pertaining to this relationship are likely acceptable. Stationarity is defined as stability of variance, mean, and autocorrelation identities across a time series, and can be thought of, in this case, as a statistical application of the principle of uniformitarianism.

Relating the results of these independent studies relies on a series of assumptions, which introduce uncertainty. Epstein et al. (1953) and Grossman and Ku (1986) corrected for the $\delta^{18}O$ of seawater, shown in Equation 8 as δ_w. Time series data sets of δ_w do not exist, so I could not perform correction for δ_w. Instead, it is assumed that the $\delta^{18}O$ of the water surrounding a carbonate as it precipitates introduces negligible or random variance in coral $\delta^{18}O$. Additionally, using this equation developed from work by Epstein et al. (1953):

$$\delta^{18}O_c = h_n * (T_{calc}) + \delta_w \quad (9)$$

where $\delta^{18}O_c$ is the $\delta^{18}O$ of a carbonate, T_{calc} is the calcification temperature of the carbonate, h_n is the linear coefficient relating the two, and δ_w is the $\delta^{18}O$ of the water surrounding the carbonate, I can relate δ_w to e_1 as drawn from Equation 1. In work discussed later in this section, I constrained the relationship between e_1 and e_3, in turn constraining part of the relationship between δ_w and $\delta^{18}O$.

Linear Dependence of Sr/Ca Ratio on T:

A study by Kinsman and Holland (1969) showed that the Sr/Ca ratio in aragonite inorganically precipitated from seawater was approximately linearly dependent on T_c. Kinsman and Holland (1969) compared the ratio of Sr to Ca in aragonite to the ratio of Sr to Ca in the solution from which it was precipitated as a function of temperature through a manipulative experiment involving the precipitation from seawater of inorganic aragonite at certain temperatures. These two ratios are linked by the distribution coefficient of the precipitation reaction, which Kinsman and Holland (1969) argued is linearly dependent on temperature, through the competition of two temperature-dependent sub ratios: the ratio of the solubility

product of Sr to that of Ca, and the ratio of the activity coefficient of Sr to that of Ca. The distribution coefficient is also affected by the changing mineralogy and structure of aragonite crystals as they grow, which introduces a measure of uncertainty (Kinsman and Holland 1969). Accounting for these variables, they were able to plot a chart of the distribution coefficients for Sr in precipitated aragonite over a range of temperatures, from which, this rough linear function can be derived:

$$k_{Sr}^{A} = (-0.45)*(T) \quad (10)$$

where k_{Sr}^{A} is the distribution coefficient between aragonite and seawater. The distribution coefficient is then related to the ratio of Sr to Ca in aragonite through:

$$(Sr/Ca)_A = (k_{Sr}^{A}) *(Sr/Ca)_L \quad (11)$$

where $(Sr/Ca)_A$ is the aragonite Sr/Ca ratio, and $(Sr/Ca)_L$ is the Sr/Ca ratio of the precipitating liquid. So, Kinsman and Holland (1969) found that as temperature increases k_{Sr}^{A} decreases, which leads in turn to a decrease in $(Sr/Ca)_A$. These relationships can be manipulated into a simple linear function linking $(Sr/Ca)_A$ to T, with h = -0.09 +/- 0.03 (2σ).

Compiling subsequent estimates of the regression of coral Sr/Ca ratio on T, Gagan et al. (2000) reported an average regression:

$$Sr/Ca = (-0.062 +/- 0.014 [2\sigma])*(T) \quad (12)$$

The results of this study (Equation 5; Figure 4; Table 1; Table A2) are not independent of those summarized by Gagan et al. (2000), but report results from an additional 23 regressions (Figure 4; Table A2), for which 6 were statistically significant. My results are independent of the results of Kinsman and Holland (1969), both in data and in experimental design. However, I find

that all three studies report regression coefficients that agree within uncertainty. This suggests that aragonites formed by different species in varying environments and conditions demonstrate similar trends between their Sr/Ca ratio and the temperatures at which they precipitated, meaning assumptions of stationarity for this chemical relationship are also sound.

Linear Dependence of $\delta^{18}O_r$ (e_1) on P_r (e_3):

There is little previous work linking coral δ^{18}O directly and quantitatively to P, so significant models linking the two would provide novel insight into proxy-climate relationships. Based on the preferential incorporation of ^{16}O relative to ^{18}O in atmospheric water (Galewsky et al. 2016), it is expected that higher precipitation would lead to a reduced δ^{18}O signature in the seawater surrounding corals, in turn reducing the δ^{18}O signature of the corals themselves (Brown et al. 2006). Considerable uncertainty is involved in making the transition between the δ^{18}O of P and that of seawater, as multiple other variables, like zonal and gradient currents, affect the δ^{18}O of seawater. However, this study, accounting for collinearity between P and T (Equation 6; Figure 5; Table 1; Table A3), found 11 statistically significant data models linearly connecting coral δ^{18}O and P (Equation 7; Figure 6; Table 1; Table A4).

The median regression of $\delta^{18}O_r$ on P_r is negative, which is consistent with previous studies that predict greater P will deplete seawater δ^{18}O, in turn depleting the coral δ^{18}O (Thompson et al. 2011). Additionally, since the theoretical relationship between P and coral δ^{18}O relies on the δ^{18}O of the seawater, the significant models I derived account for some of the variance between the two isotopic ratios.

Data Assimilation Viability:

Preliminary DA of coral $\delta^{18}O$ and Sr/Ca observations into the ZC model is underway, using the results of data modeling as described by this study. The DA is being performed for the interannual timescale variations which were most directly modeled here, and best simulated by the ZC model. In these experiments, the data models were redeveloped using data for the interval 1920-2000, and the DA performed for the period 1871-2000. This DA procedure acts as a validation for the data models developed here. Data model validation requires that the data models were trained on a certain time range, and DA products based on that training for that time range were compared against DA products for a different time range, for which the data models were not trained.

Considering this study's data models' agreement with scientific understanding of stable isotope and minor element geochemistry, as discussed above, the skill of the DA product should be about the same for the interval 1920-2000 as it is for the period 1871-1919.

Figure 9 displays results from the validation experiment. The top panel of the figure describes the correlation between DA T products which used data models trained on the years 1920-2000 and DA products for the same period. The bottom panel of the figure displays a correlation map between DA T products which used data models trained on the years 1920-2000 and DA products for the years 1871-1920. These time ranges were selected to potentially mesh with validation procedures performed for PHYDA by Steiger et al. (2018).

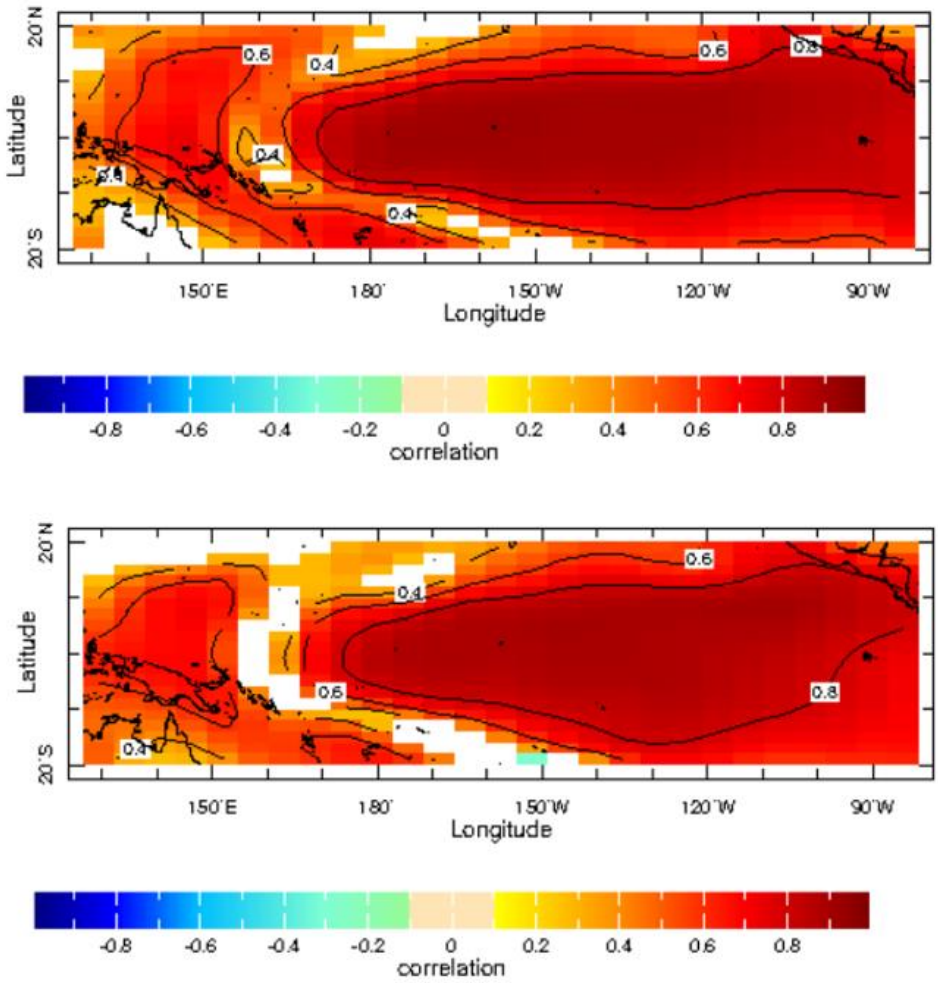

Figure (9): (Top): Correlation map between DA-estimated T with data models trained on 1920-2000 and T observations from Kaplan et al. (1998,2000) for the same period. (Bottom): As for top panel, except correlation calculated 1871-1919. Correlations are masked for significance at the $p<0.05$ level.

Differences between the two figure panels represent differences in correlation between DA products based on the ability of the data models to accurately simulate conditions during periods for which they have not been trained. Considering the similarity between the two figure panels, both in correlation amplitude and contour shape, I assert that the data models developed here are reliable for untrained time ranges. Assumptions of stationarity are further supported.

In their current state, the data models linking $\delta^{18}O$ and P are not usable in DA, as the ZC model does not explicitly simulate P. However, P can be linked to AH, which the ZC does model, through the latent heat of water (Chiang et al. 2001, Tao et al. 2016). AH as modelled by the ZC model is separated into two components, Q0o, the portion of AH reliant on T, and QF,

the total AH (Zebiak and Cane 1987). Further regressions between P and QF-Q0o, which represents the portion of AH reliant on latent heat, may constrain the DA viability of a data model linking coral $\delta^{18}O$ to AH, and are the subject of ongoing work.

Conclusions:

Least squares linear regressions between gridded, paired time series data sets of coral $\delta^{18}O$ and T, and coral Sr/Ca ratio and T, yield significant data models, which can be used through DA to update prior model states based on proxy observations. The agreement between these results and the results of prior, independent studies, points toward their stationarity and consistency with prior scientific understanding of the $\delta^{18}O$ and Sr/Ca dependencies on calcification temperature in inorganic and biogenic calcite and aragonite. Data models linking the $\delta^{18}O_r$ on P_r were also developed, which in further studies will be explored to assimilate residual coral $\delta^{18}O$ as an indicator of ZC-simulated atmospheric latent heating.

Acknowledgements:

I would like to acknowledge Justus McMillan for his constant support, my sister for reading drafts well into the night, Dr. Piccoli for his understanding, and of course my advisor, Dr. Evans, for his saintly patience.

Works Cited

Baker, B., et al. (1 Dec. 1995). "The Quality Control of Long-Term Climatological Data Using Objective Data Analysis." *Journal of Applied Meteorology and Climatology*, doi:https://doi.org/10.1175/1520-0450(1995)034<2787:TQCOLT>2.0.CO;2.

Brown, J., et al. (2006). "Modeling δ18O in Tropical Precipitation and the Surface Ocean for Present-Day Climate." *Journal of Geophysical Research*, vol. 111, no. D5, doi:10.1029/2004jd005611.

Charlson, R. J., et al. (24 Jan. 1992). "Climate Forcing by Anthropogenic Aerosols." *Science*, vol. 225, no. 5043, doi:10.1126/science.255.5043.423.

Chen, D., et al. (15 Aug. 2000). "Bias Correction of an Ocean-Atmosphere Coupled Model ." *Geophysical Research Letters*, vol. 27, no. 16, doi:10.1029/1999GL011078.

Chen, D., et al. (May 2004). "Predictability of El Niño over the past 148 years." *Nature*, doi: 10.1038/nature0243.

Chen, D., et al. (Jan. 2008). "Coupled DA For Enso Prediction." *Advances in Geosciences - A 6-Volume Set Volume 18: Ocean Science (OS)*, pp. 45–62., doi:10.1142/9789812838148_0003.

Chiang, J., et al. (2001). "Relative Roles of Elevated Heating and Surface Temperature Gradients in Driving Anomalous Surface Winds Over Tropical Oceans." *Journal of the Atmospheric Sciences*, vol. 58, no. 11, pp. 1371–1394., doi:10.1175/1520-0469(2001)058<1371:rroeha>2.0.co;2.

Cobb, K., et al. (4 Jan. 2013). "Highly Variable El Niño–Southern Oscillation Throughout the Holocene." *Science*, vol. 339, no. 6115, doi:10.1126/science.1228246 .

Collins, M., et al. (2010). "The Impact of Global Warming on the Tropical Pacific Ocean and El Niño." *Nature Geoscience*, vol. 3, no. 6, pp. 391–397., doi:10.1038/ngeo868.

Emile-Geay, Julien., et al. (2017). "A Global Multiproxy Database for Temperature Reconstructions of the Common Era." *Scientific Data*, vol. 4, no. 1, doi:10.1038/sdata.2017.88.

Epstein, S., et al. (Nov. 1953). “Revised Carbonate-Water Isotopic Temperature Scale.” *Geological Society of America Bulletin*, vol. 64, no. 11, p. 1315., doi:10.1130/0016-7606(1953)64[1315:rcits]2.0.co;2.

Evans, M.N., et al. (2013). “Applications of Proxy System Modeling in High Resolution Paleoclimatology.” *Quaternary Science Reviews*, vol. 76, pp. 16–28., doi:10.1016/j.quascirev.2013.05.024.

Gagan, M., et al. (2000). “New Views of Tropical Paleoclimates from Corals.” *Quaternary Science Reviews*, vol. 19, no. 1-5, pp. 45–64., doi:10.1016/s0277-3791(99)00054-2.

Galewsky, J., et al. (2016). “Stable Isotopes in Atmospheric Water Vapor and Applications to the Hydrologic Cycle.” *Reviews of Geophysics*, vol. 54, no. 4, pp. 809–865., doi:10.1002/2015rg000512.

Grossman, E., and Ku, T. (1986). “Oxygen and Carbon Isotope Fractionation in Biogenic Aragonite: Temperature Effects.” *Chemical Geology: Isotope Geoscience Section*, vol. 59, pp. 59–74., doi:10.1016/0168-9622(86)90057-6.

Kaplan, A. (1998-2000). “Analyses of Global Sea Surface Temperature 1856–1991.” vol. 103, no. C9, pp. 18567–18590., doi:10.1029/98jc01736.

Kaplan, A. (2015-2020). “Data Library Kaplan Extended.” *IRI/LDEO Climate Data Library*, https://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED/index.html

Kaplan, A. “USER'S GUIDE FOR THE ZEBIAK/CANE COUPLED MODEL.” *Climate Modeling and Diagnostics Group Home Page*, rainbow.ldeo.columbia.edu/~alexeyk/ZC/ZCdoc.

Kinsman, D., and Holland, J. (1969). “The Co-Precipitation of Cations with CaCO3—IV. The Co-Precipitation of Sr2+ with Aragonite between 16° and 96°C.” *Geochimica Et Cosmochimica Acta*, vol. 33, no. 1, pp. 1–17., doi:10.1016/0016-7037(69)90089-1.

McCrea, J. M. (1950). “On the Isotopic Chemistry of Carbonates and a Paleotemperature Scale.” *The Journal of Chemical Physics*, vol. 18, no. 6, pp. 849–857., doi:10.1063/1.1747785.

O'Neil, J., et al. (1969). “Oxygen Isotope Fractionation in Divalent Metal Carbonates.” *The Journal of Chemical Physics*, vol. 51, no. 12, pp. 5547–5558., doi:10.1063/1.1671982.

Steiger, N., et al. (2018). “A Reconstruction of Global Hydroclimate and Dynamical Variables over the Common Era.” *Scientific Data*, vol. 5, no. 1, doi:10.1038/sdata.2018.86.

Tao, W., et al. (2016). “The Relationship between Latent Heating, Vertical Velocity, and Precipitation Processes: The Impact of Aerosols on Precipitation in Organized Deep Convective Systems.” *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 11, pp. 6299–6320., doi:10.1002/2015jd024267.

Zebiak, S., and Cane, M. (1987). “A Model El Niño–Southern Oscillation.” *Monthly Weather Review*, vol. 115, no. 10, pp. 2262–2278., doi:10.1175/1520-0493(1987)115<2262:ameno>2.0.co;2.

Latitude	Longitude	h
-10.6907	152.8115	NaN
1	172	-0.207029
3.073056	172.7469	-0.214458
-17.5	210.1667	-0.167551
-5.22	145.82	-0.338712
1.884166	202.6732	NaN
-21.2378	200.1722	-0.076936
25	279.4	0.0493068
-16.7167	146.0333	-0.173897
-16.7167	146.0333	-0.173897
-16.8167	179.2333	-0.178374
-15	166.99	0.0368877
-22.48	166.47	-0.214307
27.10583	142.1922	-0.119569
-10.6907	152.8115	NaN
-1.5	124.833	-0.278106
-22.48	166.47	-0.216415
5.87	197.87	-0.276333
-22.1	153	-0.07646
24.93	279.25	-0.094289
-4.1916	151.9772	-0.14926
27.1059	142.1941	-0.119555
25.38	279.83	0.0589614
7.997786	277.9707	-0.125382
-22.48	166.47	-0.237097
-0.53296	166.9283	-0.651772
-5.217	145.817	-0.222245
-16.8167	179.2333	-0.239725
-5.2167	145.8167	-0.222246
1.63	124.833	-0.346968
-0.535581	166.9655	-0.344871
-16.8167	179.2333	-0.112053

Table (A1): A table displaying the latitude, longitude, and linear coefficient yielded by each regression of coral $\delta^{18}O$ on T. Significant data are highlighted in pale orange. The median slope coefficient with confidence interval is at the bottom in bright orange.

1.387978	173.025	-0.205434
-17.5	210.17	-0.178084
-0.4084	268.766	-0.12676
-21.2378	200.1722	-0.112644
7.2708	134.3837	-0.219677
7.2708	134.3837	-0.219677
-16.82	179.23	-0.321319
25.38	279.83	0.0589614
-21.2378	200.1722	-0.090029
-3.411787	143.637	-0.315818
7.2859	134.2503	-0.186139
-17.73	148.43	-0.29825
-0.4	268.77	-0.127948
-15.94	166.04	-0.281948
7.2859	134.2503	-0.223207
-5.22	145.82	-0.338712
1.884166	202.6732	-0.186886
1	173	-0.275776
13.5982	144.8359	-0.216524
-22.48	166.45	-0.302566
-15.94	166.04	-0.281717
10.2773	250.7869	-0.247777
-15.7	167.2	-0.249264
-17.73	148.43	-0.29825
-4.15	144.8833	-0.378013
-4.15	144.883	-0.37801
-21.2378	200.1722	-0.090029
7.95	278	-0.125089
10.2773	250.7869	-0.068787
-4.18	151.98	-0.14923
13.598	144.836	-0.216525
-16.82	179.23	-0.278366
-22.48	166.47	-0.214792
-15.61691	167.1311	-0.250989
3.906228	200.7141	-0.187909
1.866667	202.6	NaN
0.933	173	-0.276233
-16.8167	179.2333	-0.112053
-22.09556	152.4936	0.1114062
Median h	**-0.22 (-0.32, -0.14)**	

Latitude	Longitude	h
-16.7167	146.0333	-0.05675
24.6	277.7	-0.0084
5.87	197.87	-0.07079
-16.82	179.23	-0.05044
-16.8167	179.2333	-0.02007
-10.6907	152.8115	NaN
-4.18	151.98	0.016221
-19.93333	185.2833	0.006136
3.073056	172.7469	-0.07429
10.2773	250.7869	-0.07621
-17.73	148.43	-0.04483
-4.1916	151.9772	0.016229
-16.82	179.23	-0.06004
27.106	142.194	-0.05044
-21.2378	200.1722	-0.0316
27.1059	142.1941	-0.02559
-2.5	150.5	-0.05692
27.10583	142.1922	-0.02559
-18.315	146.595	-2.41E-05
-16.8167	179.2333	-0.02088
-15.61691	167.1311	-0.0402
10.2773	250.7869	-0.08495
-17.73	148.43	-0.04483
5.87	197.87	-0.07079
-22.475	166.4667	-0.05176
10.2773	250.7869	-0.06322
-19.9333	185.2833	0.006136
-10.6907	152.8115	NaN
-2.7	151.05	-5.48E-05

Table (A2): A table displaying the latitude, longitude, and linear coefficient yielded by each regression of coral Sr/Ca ratio on T. Significant data are highlighted in pale orange. The median slope coefficient with confidence interval is at the bottom in bright orange.

-15.7	167.2	-0.04016
-16.7167	146.0333	-0.05675
10.2773	250.7869	-0.09146
Median h	**-0.06 (-0.09, -0.02)**	

Latitude	Longitude	h
-10.6907	152.8115	101.6908
1	172	76.53678
3.073056	172.7469	0.947089
-17.5	210.1667	11.19574
-5.22	145.82	56.36996
5.87	197.87	45.81194
1.884166	202.6732	10.87369
-21.2378	200.1722	87.55439
25	279.4	45.81194
-16.7167	146.0333	56.37
-16.8167	179.2333	-33.4553
-16.8167	179.2333	10.87183
-15	166.99	-24.5683
-22.48	166.47	45.81194
27.10583	142.1922	106.9214
-10.6907	152.8115	11.20407
-10.6907	152.8115	56.36996
5.87	197.87	11.20488
-4.18	151.98	88.65487
-22.1	153	107.2765
3.073056	172.7469	56.36996
24.93	279.25	71.86507
27.10583	142.1922	0.955449
10.2773	250.7869	10.59037
-4.1916	151.9772	21.02014
27.1059	142.1941	21.02014
25.38	279.83	56.3768
-17.73	148.43	16.00771
7.997786	277.9707	21.0047
-22.48	166.47	28.9825
-0.53296	166.9283	10.59691
-5.217	145.817	41.27492

Table (A3): A table displaying the latitude, longitude, and linear coefficient yielded by each regression of P on T. Significant data are highlighted in pale orange. The median slope coefficient with confidence interval is at the bottom in bright orange.

-16.8167	179.2333	11.19574
20.83	273.26	NaN
27.106	142.194	76.97103
-21.2378	200.1722	-46.2502
-5.2167	145.8167	45.75541
1.63	124.833	41.27492
-2.5	150.5	NaN
-0.535581	166.9655	45.99476
-16.8167	179.2333	28.9825
-17.73	148.43	12.38439
1.387978	173.025	12.38412
-17.5	210.17	-24.6507
-21.2378	200.1722	-33.4451
7.2708	134.3837	-46.2496
7.2708	134.3837	56.3768
25.38	279.83	45.81194
-18.315	146.595	55.48394
-16.8167	179.2333	106.3495
-3.411787	143.637	77.88667
7.2859	134.2503	56.36996
Median h	**56.37 (24.33, 108.32)**	

Latitude	Longitude	h
-10.6907	152.8115	-0.00074
1	172	-0.0004
-16.7167	146.0333	-0.00034
24.6	277.7	-0.00092
5.87	197.87	-0.00161
1.884166	202.6732	0.000304
25	279.4	-0.00016
-16.82	179.23	-0.00113
-16.7167	146.0333	-0.00083
-16.8167	179.2333	-0.00206
-15	166.99	-0.00075
27.10583	142.1922	-0.00166
-10.6907	152.8115	-0.00211
-1.5	124.833	7.64E-05
-10.6907	152.8115	-0.00094

Table (A4): A table displaying the latitude, longitude, and linear coefficient yielded by each regression of the $\delta^{18}O_r$ on P_r. Significant data are highlighted in pale orange. The median slope coefficient with confidence interval is at the bottom in bright orange.

-22.1	153	-0.00082
27.10583	142.1922	-0.00018
-17.73	148.43	-0.00114
-22.48	166.47	-0.00083
-5.217	145.817	-0.00025
-16.82	179.23	-0.00018
27.106	142.194	-0.00113
-21.2378	200.1722	-0.00016
1.63	124.833	-0.00028
27.1059	142.1941	-0.00045
-0.535581	166.9655	-0.00113
-16.8167	179.2333	-0.00115
-17.73	148.43	7.64E-05
Median	**-0.00011 (-0.0019, -0.00044)**	