Spring 4-23-2021

Diversity – Independent Factors Predict Elevated Extinction Rates

Dustin Perriguey

Corinne Myers
University of New Mexico

Jason Moore
University of New Mexico

Louis Scuderi
University of New Mexico

Recommended Citation

Perriguey, Dustin; Corinne Myers; Jason Moore; and Louis Scuderi. "Diversity – Independent Factors Predict Elevated Extinction Rates." (2021). https://digitalrepository.unm.edu/eps_etds/319

Dustin Perriguey

Candidate

Earth and Planetary Sciences

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Corinne Myers , Chairperson

Louis Scuderi

Jason Moore

DIVERSITY – INDEPENDENT ENVIRONMENTAL FACTORS PREDICT ELEVATED EXTINCTION RATES

by

Dustin Perriguey
BS Earth and Planetary Sciences, University of New Mexico, 2018

Department of Earth and Planetary Sciences
University of New Mexico

Master of Science in Earth and Planetary Sciences
May, 2021

ACKNOWLEDGEMENTS

I would like to thank my family and friends for supporting me through this process, my girlfriend for convincing me to finish my degree, being incredibly supportive, and providing constructive feedback, Paula Pascetti and all the front office staff for being the rock of the Earth and Planetary Sciences Department at UNM, all the graduate and undergraduate students that contributed to the databases used in this study, Cori and Remy for all your guidance, all the authors of the published data used in this study for providing data and helpful conversation about these large-scale patterns, my committee for being thoughtful and supportive, Dr.'s Pun, Zhang, and Sharp for their guidance, the Brooklyn College Crew! for teaching me that scientists can do science while still being real people, and all the other grad students I've worked with in the EPS department.

Diversity – Independent Factors Predict Elevated Extinction Rates

By Dustin Perriguey
Associate of Arts in General Studies, Eastern New Mexico University – Ruidoso, 2015
Bachelor of Science in Earth and Planetary Sciences, University of New Mexico, 2018
Degree to be awarded: Master of Science in Earth and Planetary Sciences, University of New Mexico, 2021

ABSTRACT

Multiple linear regression was used to determine the relationships between diversity-independent factors (i.e., abiotic, climatic) 2, 5, and 10 Myrs-prior to the most elevated Phanerozoic extinctions. We constructed five abiotic variables from Phanerozoic proxy records[1–5] to compare to extinction rates: mean temperature, temperature instability, carbon cycle instability, continental weathering rates, and habitat instability. All three models were statistically significant ($P < 0.05$) and explained > 70% of the variation in Alroy's[6] three-timer generic extinction rates. However, the 2 Myr-prior model explained the most variance in extinction rates and had the most predictive power, based on adjusted and predictive R^2 (~ 72% and 41%, respectively). Carbon cycle and habitat instabilities significantly contributed to this model ($P < 0.05$), thus suggesting that these variables positively contribute to the most severe extinctions during the Phanerozoic. However, carbon cycle and habitat instabilities seem to behave as extinction intensifiers, requiring an additional trigger to set off a major extinction event. Using the equation of the best fit line of the 2 Myr-prior model and the significant variables carbon cycle and habitat instabilities, we predicted a modern three-timer generic extinction rate of 0.85 (PI: 0.29, 1.40), falling between the end-Ordovician and end-Triassic mass extinctions in taxonomic severity. These results provide important information regarding the role diversity-independent factors play in intensifying the most elevated extinctions during the Phanerozoic and will continue to play in our present and future. Furthermore, these results support the importance the fossil record for contextualizing the potential severity of the modern extinction crisis.

Table of Contents

Introduction

The marine fossil record provides a fundamental resource to inform predictions of the severity of the modern extinction crisis[7–9]. It is not only essential to use the fossil and geologic records to set baselines for conservation policy [10], it is necessary that we use the dynamic record of Earth-life interactions in the paleobiologic record to inform predictions of the consequences of anthropogenic climate change and human-driven extinction on future biodiversity. We modelled the influence of abiotic environmental conditions (e.g., temperature or sea level) on elevated extinction rates during the Phanerozoic. Our hypothesis that initial abiotic conditions have an amplifying or dampening role in the most elevated Phanerozoic (~ 541 Ma to Recent) extinctions, was evaluated using multiple linear regression models. Model output allowed for estimation of future extinction rates of marine genera under "modern" environmental conditions.

Macroevolutionary extinction hypotheses have been generated and tested at ecological and geological time scales[11–14]. Extinction and origination are the basic parameters used to estimate diversity in these models and are the drivers of diversity over geologic timescales. Extinction is regulated by the interaction of biotic and abiotic factors[15,16] (e.g. competition, predation, temperature, pH, nutrient availability, etc.). Although there are many hypotheses with varying degrees of support regarding the proximate triggers of major past extinctions, relationships between extinction severity and boundary conditions have not been investigated as heavily. Negative diversity-dependence (high diversity reducing diversification rates) has been invoked as a complex of biotic factors that contributed to the correlation between Phanerozoic extinction and preceding diversity[13,17,18]. However, these analyses did not include the role of diversity-independent factors (i.e., abiotic, climatic) in shaping extinction susceptibility. However, diversity-independent factors have been found to be strong predictors of diversity, extinction, and origination throughout the Phanerozoic[19–21].

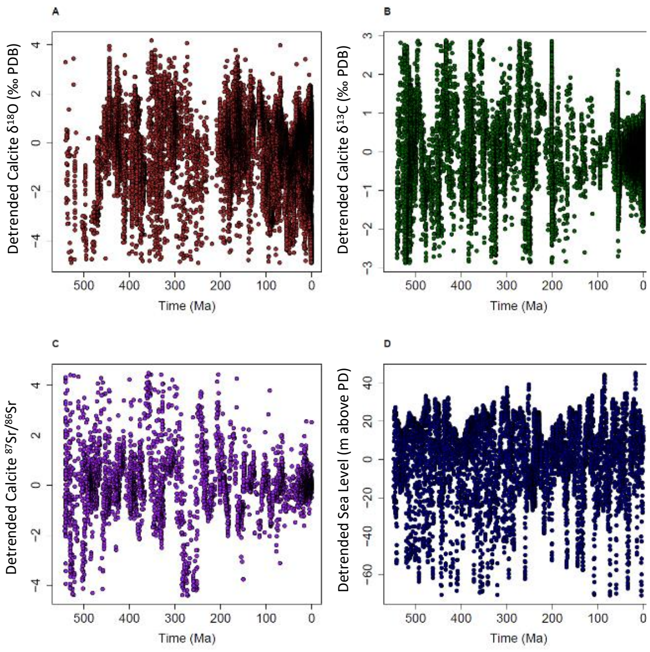
Published Phanerozoic extinction[6,22] and abiotic proxy records[1–5] were used to determine if certain Earth system states preconditioned high or low extinction rates once an extinction event was triggered. Two different extinction rate estimates were used to identify sixteen common periods of elevated Phanerozoic extinction: Alroy's[6] three-timer generic extinction and Bambach's[22] re-analysis of the original Sepkoski Phanerozoic diversity compendium[23]. Generic extinction rates were used in both studies because they increase the sample size and are comparable with earlier studies[24]. Although both the Alroy[6] and Bambach[22] data were used to establish the dataset of elevated Phanerozoic extinctions, Alroy's[6] extinction rates are focused on here because they better consider the many preservation and sampling biases that negatively impact a literal interpretation of the fossil record[24] (see SI for Bambach-based analytical results).

On long time scales, geochemically and geophysically sourced proxies record Earth system evolution in the form of climatic and spatial changes. Global climate is heavily influenced by the carbon cycle, which is primarily modulated by the interaction and associated feedbacks of atmospheric $p\mathrm{CO}_2$, chemical weathering, temperature, the hydrologic cycle, and tectonics[25]. In addition to climatic-driven conditions in the form of temperature, ocean oxygenation, and ocean acidification, marine taxonomic relationships with area of available habitat (i.e., habitat loss) has been found to be a driver of extirpation and biodiversity loss in modern settings[26,27]. Thus our analysis focused on abiotic environmental records of oxygen (δ^{18}O), carbon (δ^{13}C), and strontium (^{87}Sr/^{86}Sr), as well as sedimentological eustatic sea level (SL)[1–4] to approximate climate and habitat conditions. We derived six abiotic proxies from these records: mean temperature ($\bar{x}$ δ^{18}O), temperature instability (s δ^{18}O), carbon cycle instability (s δ^{13}C), continental weathering rates (m ^{87}Sr/^{86}Sr), and habitat instability (s SL) where $\bar{x}$, s, and m are mean, standard deviation, and slope, respectively. These environmental records and proxies were subset to 2, 5, and 10 Myr bins prior to episodes of elevated extinction to further test for temporal sensitivity.

Multiple linear regression was used to test the hypothesis that preceding environmental factors (i.e., $\bar{x}$ δ^{18}O, s δ^{18}O, s δ^{13}C, m ^{87}Sr/^{86}Sr, and s SL) explain the variation in the most elevated Phanerozoic three-timer extinction rates. Extinction rates were then estimated for the modern extinction crisis using the results from the regression model with the most explanatory power while accounting for model complexity. This approach demonstrates the utility of the fossil and geologic records to inform predictions about impending extinction vulnerability.

Results

Multiple linear regression models were generated at different temporal bin sizes using the R Programming environment[28]; these sensitivity analyses demonstrated that the 2 Myr-prior model was the least overfit and had the most predictive power (Table 1 and SI for other model results). The ^{87}Sr/^{86}Sr data were multiplied by 10^4 to allow for conceptually more comparable slopes in the linear models. All abiotic environmental proxies were detrended using LOESS smoothing to remove secular trends and focus on the sub-trends within these data (see SI). After detrending, outliers in the environmental proxy datasets were identified as observations outside three standard deviations of the datasets, and were removed prior to modeling given the high likelihood that these proxies had modified or reset values due to



	P-Value	Multiple R^2[ⱡ]	Adjusted R^2	Predicted R^2
2 Myr-prior Model[*]	0.010	0.814	0.721	0.409
Mean Temperature	0.893	0.026		
Temperature Instability	0.318	0.016		
Carbon Cycle Instability	0.000	0.515		
Continental Weathering Rates	0.514	0.037		
Habitat Instability	0.018	0.221		

Note:

[*] Bonferroni corrected P-Value

[ⱡ] Subsequent values are relative R^2

Fig. 1 | Detrended environmental proxy data and select 2 Myr-prior model fitness indices. Diversity-independent (i.e., abiotic, climatic) environmental proxies used here were $\delta^{18}O$[19], $\delta^{13}C$[20], $^{87}Sr/^{86}Sr$[18], and eustatic SL[21], sourced from and supplemented with the primary literature[19,22]. Abiotic environmental proxy data were subjected to LOWESS regression, a non-parametric, polynomial regression, through which the residuals can be calculated from the fitted polynomial (top left). These residuals are extracted and then used in place of the original data. This removes the secular trends and allows for analysis of sub-trends in the data. We constructed five abiotic environmental variables from these data ($\bar{x}$ $\delta^{18}O$, *s* $\delta^{18}O$, *s* $\delta^{13}C$, *m* $^{87}Sr/^{86}Sr$, and *s* SL) and used multiple linear regression to model their effects on the most elevated Phanerozoic three-timer extinction rates.

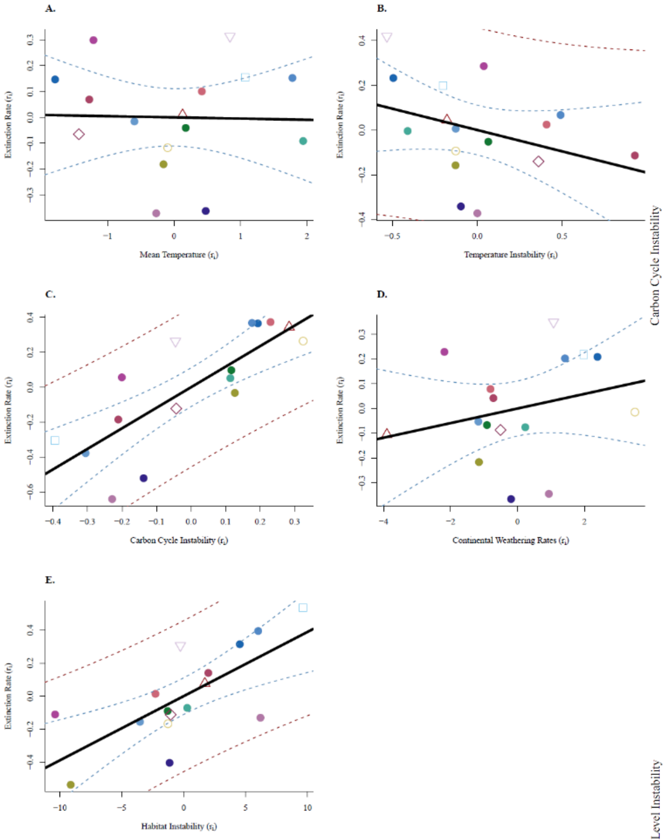
Table 1 contains model fitness indices of the 2 Myr-prior multiple linear regression analysis between three-timer elevated extinction rates and variables constructed from the detrended proxy data (bottom left). These model fitness indices indicate that this statistically significant model explained ~ 81% of the variance in extinction rates and ~ 72% when accounting for the number of predictors in the model. Also, the model output indicated carbon cycle and habitat instabilities as having had significantly contributed to the model. Partial R^2 of the predictors reveals carbon cycle and habitat instabilities as explaining the most variation in the most elevated three-timer extinction rates, ~ 52% and ~ 22%, respectively. The predicted R^2 suggests a moderately predictive, slightly overfit model. This model was determined to be the least overfit and have the most predictive power.

tectonic and diagenetic alteration[29–31]. The models did not grossly violate the assumptions of multiple linear regression[32] (see SI).

The 2, 5, and 10 Myr pre-extinction models explained > 70% of the variance in three-timer extinction rates (see SI). However, using a leave-one-out cross-validation technique (predictive R^2) indicated overfit models for all models with the 2 Myr-prior model classified as least overfit[33]. The 2 Myr-prior model had an adjusted R^2 of ~ 72% and had a predictive R^2 of ~ 41%, indicating a moderately predictive, but slightly over fit model. Only carbon cycle and habitat instability variables were statistically significant in the model and explained 51.5% and 22.1% of the variation, respectively (Table 1). These results indicate a strong correlation between elevated Phanerozoic extinction rates, carbon cycle instability, and habitat instability.

The possibility of no relationship was detected between three-timer extinction rates and mean temperature, temperature instability, and continental weathering rates using the 95% confidence bands in the predictor effect plots (PEP, Fig. 2 A, B, and D). In contrast, the 95% confidence band for carbon cycle and habitat instabilities does not include the possibility of no relationship with extinction rates (Fig. 2 C and E). This suggests that most severe extinctions occurred after two-million-year intervals of increased carbon cycle and habitat instabilities.

PEP for each of the variables and three-timer extinction rates reveal that these trends are primarily driven by extreme abiotic parameter values and extinction rates. The highest and shared abiotic trends are occurring during the Stage 3 – Stage 4, Series 3 – Furogian, and end-Permian intervals followed by three-timer extinction rates at ~ 1.41, 1.50, and 1.79, respectively, against the mean of 0.68 for all events. We find all models to be poor estimators of extinction rates when these are removed (see SI). This contrasts with the same process applied to the relationship between current diversity and future extinction rates found by Alroy[13]. However, these are well-known and well-characterized intervals of elevated extinction that describe important periods of biodiversity decline in the Phanerozoic. Therefore, even though they are very extreme, removing them would prevent a realistic evaluation of the hypothesis that initial abiotic environmental conditions contribute to the enhancement of elevated extinctions[22,34]. We further assessed for the effect of environmental variable sampling intensity (i.e., number of datapoints in a bin) on extinction rates and found no significant relationships, except for $\delta^{18}O$ (Spearman rank-order correlation $\rho = -0.645$, P =



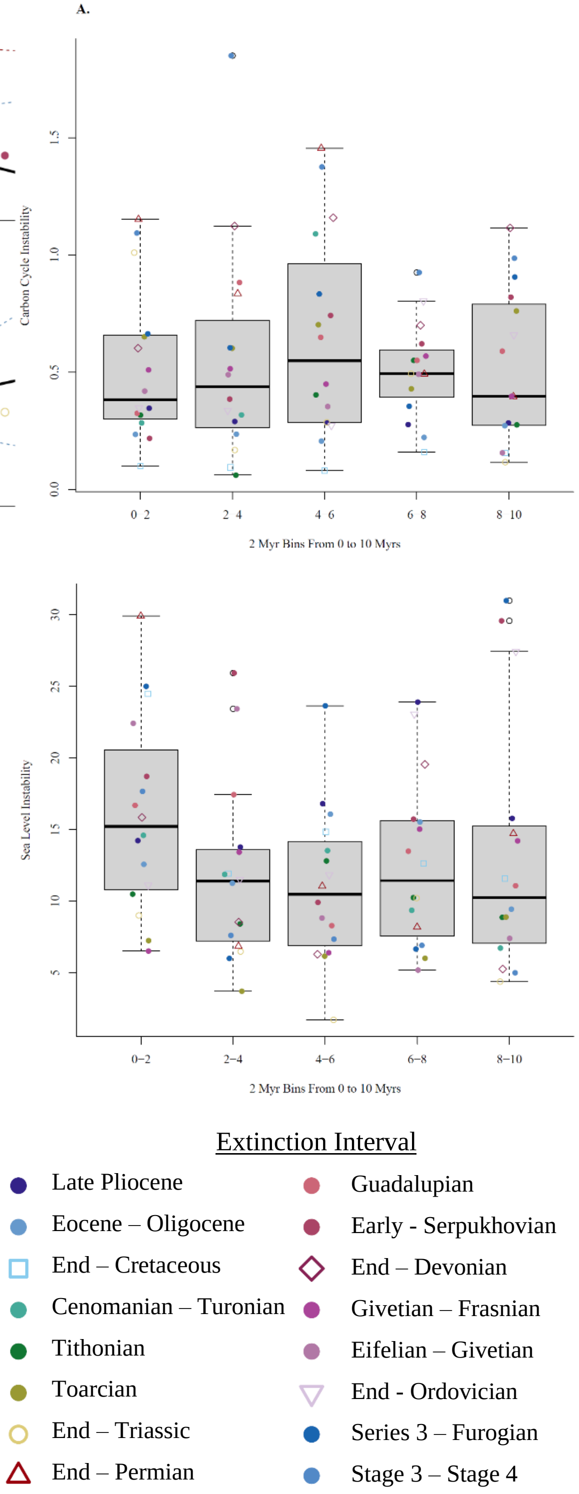


Fig. 2 | Partialled out effects plots of the 2 Myr-prior model and 2 Myr stepped boxplots of carbon cycle and habitat instability.

The 2 Myr-prior model contained five predictor variables ($\bar{x}$ δ^{18}O, *s* δ^{18}O, *s* δ^{13}C, *m* ^{87}Sr/^{86}Sr, and *s* SL). Using partialled out effects plots (above), also called partial residual plots, we can see the effects of individual variables while adjusting for the effects of all other variables in the model. The plots contain the 95% CI (inner dotted lines) and PI (outer dotted lines). Note the only viable trends between three-timer extinction rates and carbon cycle or habitat instabilities.

Boxplots were constructed (right) to assess for the possibility of these two significant variables behaving as extinction triggers. Carbon cycle and habitat instabilities were binned into 2 Myr intervals from 2 Myrs up to 10 Myrs. If these did behave as triggers, we would expect them to depict an exponential decay. However, they do not and thus only behave as extinction intensifiers.

0.007). This provides evidence that the 2 Myr-prior *s* δ^{13}C and *s* SL effect on three-timer extinction rates is a real signal. See SI for tables 2 and 3 containing the number of abiotic variable datapoints per bin and their summary statistics.

Whereas the 2 Myr model was classified as the least overfit model, models binning data over longer time intervals (i.e., 5 and 10 Myr-prior) were statistically significant at $P < 0.05$ and shared carbon cycle instability as a statistically significant predictor variable $P < 0.05$ (see SI). This larger pattern demonstrates the strength of preceding carbon cycle instability as a predictor of subsequent elevated extinction rates. To investigate this pattern further, we constructed boxplots of carbon cycle and habitat instability into 2 Myr intervals from 2 Myrs up to 10 Myrs prior to elevated extinction intervals (Fig. 2). If carbon cycle or habitat instability directly triggered these elevated extinctions, we would expect to see the 2 Myr bins resembling an exponential decay function, otherwise these parameters behave solely as extinction intensifiers. This exercise was used to determine the possibility of these environmental variables contributing to triggering the extinction events. These boxplots demonstrate no clear trend, which suggests that although the carbon cycle and habitat instability factors likely worsened an extinction event, a different and/or larger perturbation was required to trigger episodes of elevated extinction rates. (see SI for further analyses).

The results of our models provide evidence in support of carbon cycle and associated global climate flux 2 Myrs prior to the most elevated three-timer Phanerozoic extinction rates. Carbon cycle and habitat instabilities significantly contributed to the 2 Myr-prior model, the least overfit model explaining the most variance of three-timer extinction rates. Both *s* δ^{13}C and *s* SL were positively correlated with extinction rates and these results can be explained by the fact that the burial of carbon in the ocean is modulated by sea level[3]. These results suggest that prolonged periods of biotic stress, caused by the combined effect of carbon cycle and habitat fluctuations, may have “primed” taxa toward destabilization before encountering a larger environmental perturbation that triggered extinction. In other words, highly volatile Earth system conditions may predispose species to be less successful at weathering additional environmental stress from a large igneous province, asteroid impact, rapid climate change, or other major extinction trigger. This indicates that, prior to extinction events, diversity-independent environmental factors may influence subsequent extinction severity significantly through time, and may help explain the unexplained variance in negative diversity-dependent models[13,17,18].

Determining how interannual and longer, large-scale patterns of climate variability (e.g., North Atlantic and El Niño-Southern Oscillations) impact modern ecological structure is becoming a subject of more recent interest to population ecologists[35–37]. These studies apply the concepts of the so-called Moran effect, or synchronous population fluctuations mediated by abiotic fluctuations (i.e., weather or climate variability), to understand organism dispersal patterns, abundance structure, range shifts, and disruption of tightly connected trophic interactions[35,38]. Diversity-dependent and -independent factors interact to produce the ecological patterns observed in specific and interspecific populations[35,39]. A Moran effect at the geologic scale is likely at play when considering the results described here with those found regarding negative diversity-dependent diversification rates[13,17,18]. The implication being that there is a synergistic effect of diversity-dependent and -independent factors that initiated destabilization of ecological and evolutionary processes prior to the most elevated extinction events. These factors likely influence each other in complex ways and should be studied in more detail in both the modern and geologic records[35,37], through which important understanding for the modern extinction crises can be derived.

Modern Extinction Crisis

It has been argued whether modern humans (*Homo sapiens*) have had an effect on the Earth system since our evolution and dispersal[40–43]. On the other hand, anthropogenic climate change and habitat fragmentation/loss have been invoked as major drivers of the modern extinction crisis[44–46]. Our models provide a potential test of the sensitivity of our current Earth system to the types of anthropogenic triggers that may provoke elevated extinction rates in our near future. Using our results above, we can ask a simple question: has the Earth system over the last 2 Myrs been in a state through which the modern extinction crisis could become further accelerated?

Using the significant abiotic environmental variables s δ^{13}C and s SL and the equation of the best fit line from the 2 Myr-prior model, we calculated a modern three-timer extinction rate of 0.85 (PI: 0.29, 1.40) for the "6th Mass Extinction" provided current anthropogenic activities are strong enough to trigger an interval of elevated extinction rates. This predicted rate falls in intensity between the end-Ordovician and end-Triassic three-timer extinction rates of 0.81 and 0.94, respectively. We chose to only include s δ^{13}C and s SL in the model prediction because they were the only variables whose fit for the line did not contain the possibility of no relationship (i.e., did not contain the possibility of zero slope).

Model results indicate abiotic preconditioning of the current Earth system, which has implications for the modern extinction crisis. Our estimate of a potential modern extinction rate is within the realm of some of the most severe of the "Big5" mass extinctions. This suggests that the state of the Earth's carbon cycle over the last 2 Myrs is likely exacerbating the current extinction crisis.

Notably, these results pertain to generic, not species level, extinction (the latter being the standard focus when investigating modern extinction risk). Although much previous work supports the utility of genera as proxies for species in diversity studies[47,48], the magnitude of these results and the conflicting perspectives on the modern extinction crisis[8,9] support the pressing need to improve translatable metrics between fossil and modern biodiversity records[8,49]. Nevertheless, our results provide useful information about the distal contributions to extinction and the role of the initial abiotic environment in enhancing periods of elevated extinction rates. Furthermore, the finding that high carbon cycle and sea level fluctuations had the strongest influence on elevated extinctions, lends additional support to the need to focus on interactions between the global carbon cycle, its perturbations, and consequences for the Earth's biosphere[3,50,51].

Methods

Environmental Proxy and Extinction Rate Data

Three-timer extinction rate data[6] was sourced through personal communication with J. Alroy. The environmental proxy data $\delta^{18}O$, $\delta^{13}C$, $^{87}Sr/^{86}Sr$, and eustatic SL were sourced from Veizer and Prokoph[2], Bachan et al.[3], Prokoph and Veizer[1], and Boulila[4], respectively. The $\delta^{18}O$, $\delta^{13}C$, and $^{87}Sr/^{86}Sr$ data were supplemented using Zaky et al.[5] and Veizer and Prokoph[2]. The Veizer and Prokoph[2], Bachan et al.[3], Prokoph and Veizer[1], and Boulila[4] data were all geologic timescale (GTS) standardized to GTS2012, GTS2012, GTS2004, and GTS2016, respectively. The Zaky et al.[5] data were GTS standardized to GTS2012.

Environmental Proxy Data

The environmental proxy data $\delta^{18}O$, $\delta^{13}C$, $^{87}Sr/^{86}Sr$, and eustatic SL are argued to represent temperature, carbon cycle, continental weathering, and habitat availability on geologic timescales.

Oxygen isotope values are largely influenced by temperature but can face a number of diagenetic overprints. However, Veizer and Prokoph[2] argue against the probability of worldwide increased diagenetic alteration in older sedimentary rocks in favor of an evolving ^{18}O enrichment from a decline in the planetary thermal regime, hydrothermal interactions, and a dynamic Earth system toward the Recent.

Carbon isotope values are sensitive to atmosphere-ocean carbon cycle dynamics reflected in dissolved inorganic carbon in seawater, but are also influenced by primary productivity, ocean circulation, sedimentation rate, ocean latitudinal and depth gradients, and are fairly insensitive to diagenetic alteration[52]. The $\delta^{13}C$ values were sourced from bulk carbonate rock and marine fossil calcite compilation which show clear excursions because of high resolution sampling and meet low diagenetic alteration protocols[3].

The $^{87}Sr/^{86}Sr$ is the ratio of continental fluvial input to hydrothermal mantle input[53], and because of the fluvial input we can consider this a proxy for continental weathering. Strontium isotopes are well mixed with respect to the seawater with a known residence time of approximately $\geq$ 4Myr and are independent of organism habitat or their physiological effects, therefore the $^{87}Sr/^{86}Sr$ is interpreted as a global signal[1,54]. Samples in this dataset come from fossils with low-Mg calcite, micritic carbonate (nanoplankton), and whole rock[1].

Eustatic SL were sourced from globally correlated, tectonically inactive/dormant or correctable sequence-stratigraphic studies[4].

Extinction Rate Data

Bambach[22] originally identified eighteen extinctions from the Phanerozoic that fit the Sepkoski 1986 definition of a mass extinction[55]. We originally constructed models using extinction rates calculated from both Bambach[22] and Alroy[6] to compare their performance. The Alroy[6] three-timer extinction rate data could not be subset to the Bambach[22] eighteen due to bin size differences (i.e., ~ 3 and 10 Myr, respectively), and therefore only sixteen three-timer extinction rate observations were extracted. The Bambach models performed poorly and consequently we did not include their results in the main text (see SI for results). We only consider the methods as they pertain to the three-timer extinction rate models throughout the rest of this section. However, the methods here are identical to those applied to the Bambach[22] data, with exception to any model interpretation due to lack-of-fit and of statistical significance.

The Alroy[6] three-timer extinction rate dataset was chosen for these analyses because they explicitly attempt to account for biases in fossil preservation and sampling. These data were originally downloaded from the Fossilworks website with a number of downloading criteria to account for preservation biases and sample standardized using the shareholder quorum subsampling method[56]. Alroy[6] used the following downloading criteria from Fossilworks: exclusion of Tetrapoda, Ostracoda, Arachnida, and Insecta metazoans; exclusion of terrestrial collections, collections of unlithified sediments, preserved soft parts, compression or aragonitic fossils, or fossils collected by bulk sieving, collections at the geographic scale of a basin or geological group, occurrences with genus names qualified by aff., ex gr., sensu lato, or quotation marks; occurrences of informal taxa, ichnofossils, form taxa, and occurrences unresolved to the taxonomic level of a genus; and subgenera were treated as distinct genera. The three-timer rate equation is a log normal ratio of genera sampled before, within, and after a bin, all multiplied by a conditional log normal probability of being sampled given a genus was definitely present throughout the bin[56]. This equation attempts to reduce the Signor-Lipps effect and the Pull of the Recent, while also accounting for Lazarus taxa[56].

Although biases exist in compiled paleontological data such as those used in these analyses[6,24], we searched the scientific literature to ensure these data are the best currently available.

Select Model Fitness Indices

Models with statistically significant Bonferroni corrected P-values were examined for explanatory power and goodness-of-fit using multiple R^2, adjusted R^2, and predictive R^2. Multiple R^2 and adjusted R^2 were extracted from the model output and predictive R^2 was calculated. Predictive R^2, a leave-one-out cross-validation technique also called cross-validated R^2, allows for a model with a small amount of data to be assessed for predictability[33]. PRESS, essential to calculating predictive R^2, is calculated through iterative fitting of the model, but removing a different observation each time to find the prediction error of the model[33]. Predictive R^2 is calculated as the predictive residual sum of squares divided by total sum of squares (1-(PRESS/TSS)) and this alteration is only for making PRESS more easily interpretable in terms of the model it was calculated from[57,58]. Predicted R^2 and adjusted R^2 can be thought as the lower and upper bounds of how well the predictors explain the dependent variable and predict new observations[58]. Predictive R^2 is commonly used in chemometrics and drug design as a measure of predictability in partial least squares analyses and it has been determined that values greater than 30% are suggested to be significant and negative values are possible when PRESS is larger than TSS[59–61]. However, we chose a more conservative cutoff for a reasonable predictive R^2 of greater than 60% to represent a reliably predictive model. Indeed, the difference between adjusted R^2 and predictive R^2 estimates the degree of overfitting of the model. We used the PRESS() function from the MPV package in R to calculate PRESS and TSS was extracted from the model output.

Only the 2 Myr-prior model had a balance of both explanatory power and goodness-of-fit with an adjusted R^2 of ~ 72% and predictive R^2 ~ 41% (see SI Table 4). However, this predictive R^2 value indicates an overfit model. This represents that ~ 30% of the variance was explained by an overcomplicated model.

Evaluation of Predictors in the Model

Statistically significant predictors in the 2 Myr-prior model were *s* $\delta^{13}C$ and *s* SL (P = 0.0005 and 0.02, respectively) (see model output in SI for all other models). The 2 Myr-prior model was reconstructed using an interaction term *s* $\delta^{13}C$ x *s* SL to assess for interaction effects. There were no detectable interaction effects (P = 0.27), so the interaction term was removed from the model. Based on P-values alone, *s* $\delta^{13}C$ and *s* SL are important in explaining most of the variance in the model. This is also apparent when inspecting the partialled out R^2 values of the predictors (see SI Table 4). However, this is challenged by inspection of the partialled out effects plots (PEP), the slope of the least squares best-fit line for *s* SL contains the possibility of zero slope via its wide confidence band (Fig. 4). Inspection of *s* $\delta^{13}C$ PEP revealed a tighter relationship with extinction rates and narrower confidence band, not containing the possibility of zero slope.

Assessment of Statistically Significant Predictors as Extinction Triggers

The other two models, 5 Myr and 10 Myr-prior, did not have the same level of predictability as the 2 Myr-prior model, but they did all share *s* δ^{13}C as significantly contributing to the model and explaining the most variance in extinction rates. Although not interpretable, we decided to further investigate this possible trend of carbon cycle instability becoming more important the closer the model is to the extinction interval (see SI). We took 10 Myr-prior bins of *s* δ^{13}C and sliced them into 2 Myr bins before the sixteen three-timer extinction rate bins. Boxplots of the 2 Myr bins were constructed in base R[28]. This was an exploratory procedure conducted to investigate the potential for our environmental variables to have been triggers in subsequent extinctions. The *s* SL was also investigated using the procedure described above for similar reasons, although this variable had only contributed significantly to the 2 Myr and 5 Myr-prior models.

Estimating a modern three-timer extinction rate consisted of subsetting the data 2 Myr prior to time 0. We calculated $\bar{x}$ δ^{18}O, *s* δ^{18}O, *s* δ^{13}C, *m* ^{87}Sr/^{86}Sr, and *s* SL for this bin and using the equation for the best-fit-line, predicted a modern extinction rate. This was achieved using the predict() function with our new data (i.e., the last 2 Myr) in R, which outputs a fitted value and prediction interval.

Supplemental Information

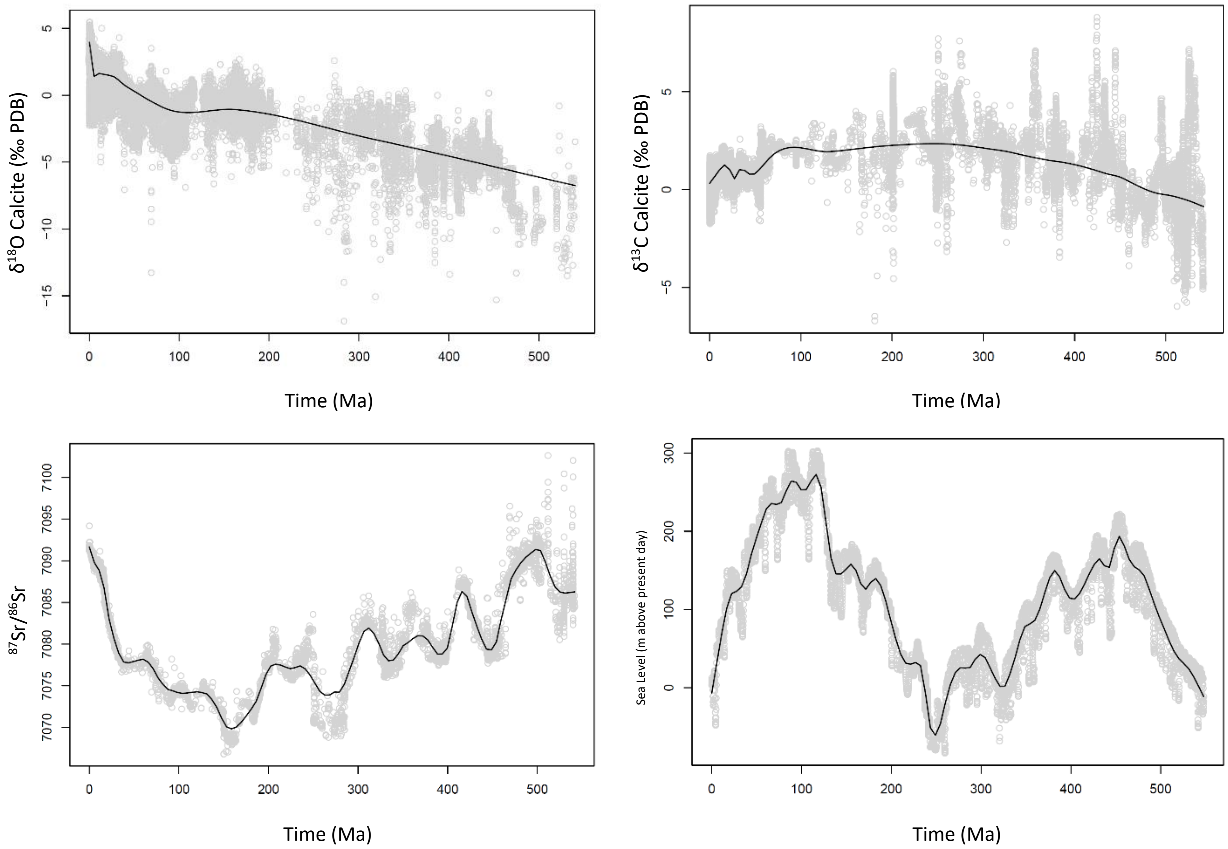


Fig. 3 | LOESS regression fitted polynomial for environmental variables. Each environmental proxy was fitted with LOESS regression. This is a non-parametric, polynomial regression, whereby the residuals can be calculated from the fitted polynomial. This effectively removes the secular trends and then the residuals are used to better assess the sub-trends in the data. See Fig. 1 to see proxies detrended.

Environmental Proxy Detrending and Outlier Removal

Proxy Detrending

All Phanerozoic extinction and environmental data used here contained some sort of secular trend. However, Alroy[6] described this trend for the three-timer extinction rate data as high rates in the Cambrian and Ordovician and low rates in the Cenozoic *appearing* to show an overall declining trend. All original Phanerozoic environmental data (stable isotopes for oxygen ($\delta^{18}O$)[1,2,5] and carbon ($\delta^{13}C$)[3,5], strontium isotope ratios ($^{87}Sr/^{86}Sr$)[1,2,5], as well as sedimentological eustatic sea level (SL)[4]) were detrended using locally estimated scatterplot smoothing (LOESS). Detrending of timeseries data can be done linearly, performing regression on the timeseries and analyzing the residuals of the least squares best-fit line. LOESS by comparison is a non-parametric, polynomial regression, whereby the residuals can be calculated from the fitted polynomial. These residuals are extracted and then used in place of the original data. We chose the lowess.as() function from the fANCOVA package in the R Programming environment[28] with automatic smoothing parameter selection (auto-span), using generalized cross-validation criterion, and fitting by re-descending M estimator outlier rejection with Tukey's biweight function to perform loess. Auto-span controls how data are weighted surrounding a fitted point but calculates the optimal span value automatically, using generalized cross-validation (GCV) methods. GCV determines a span value that minimizes the residual sum of squares of the fitted polynomial by iteratively removing points and recalculating the fitted polynomial[62]. Re-descending M estimator outlier rejection with Tukey's biweight function was used because we assumed the data had outliers. We were able to remove secular trends from all variables and all fitted polynomials seem to accurately represent the environmental timeseries' (Fig. 3). We had initially attempted to use arial extent of carbonate continental shelf, as well as other variables extracted from these studies (e.g., arial extent of clastic shelf or reef environments)[63,64]. However, their data were too coarse for these analyses (i.e., geologic eon and stage) and were not used.

Outlier Removal

Lastly, values outside three standard deviations of the data were classified as outliers and were removed from all environmental proxy datasets. Prior to removing outliers, there were 57715, 38149, 5283, and 5494 data points for $\delta^{18}O$, $\delta^{13}C$, $^{87}Sr/^{86}Sr$, and SL datasets, respectively. After removal of outliers there were 57414, 37182, 5166, and 5439 data points for $\delta^{18}O$, $\delta^{13}C$, $^{87}Sr/^{86}Sr$, and SL datasets, respectively. There was a total of 106641 before and 105201 after outlier removal, a total of 1440 data points removed.

Environmental Variable Construction

Mean, standard deviation, or slope (i.e., $\bar{x}$, *s*, and *m*) were calculated for their respective detrended variables in bins sized at 2, 5, and 10 Myr before each of the 16 extinction bins. The resulting variables were as follows: $\bar{x}$ $\delta^{18}O$, *s* $\delta^{18}O$, *s* $\delta^{13}C$, *m* $^{87}Sr/^{86}Sr$, and *s* SL. We defined these as mean temperature, temperature instability, carbon cycle instability, continental weathering rates, and habitat instability based on what their interpretations represent over long timescales[1–4]. The resulting datasets are available in tables 5 and 6.

Carbonate Continental Shelf

Areal extent of carbonate continental shelf was explored as a proxy for habitat area. Walker et. al[63] provides a useful compilation, but the data are too coarse for our analyses (i.e., at the geologic Epoch level). The use of linear interpolation between Epochs was explored to increase the resolution of these data. However, using linear interpolation between Epochs applies a gross assumption of trends at the geologic Stage or lower levels. Next the dataset compiled by Foote[64] was explored because these data have higher resolution (i.e., geologic Stage). However, here too the resolution was too coarse to match our analyses. It is likely that area of carbonate continental shelf would be a good proxy for habitat area, given t without added interpolation and associated assumptions. It is likely that area of carbonate continental shelf would be a good proxy for habitat area, given the stronger affinity marine taxa have for carbonate sediments[64], but the most recent data available are still too coarse[65]. This would be an excellent area for testing the observed relationship between habitat

Table 2: Number of Datapoints per Bin

	Oxygen Isotopes			Carbon Isotopes			Strontium Isotopes			Sea Level		
	2 Myr	5 Myr	10 Myr	2 Myr	5 Myr	10 Myr	2 Myr	5 Myr	10 Myr	2 Myr	5 Myr	10 Myr
Late Pliocene	680	1983	4603	509	1678	3696	30	159	380	20	49	100
Eocene-Oligocene	345	709	1151	183	393	737	4	12	18	24	60	111
End-Cretaceous	460	743	992	12	30	49	13	50	62	21	53	107
Cenomanian-Turonian	169	301	307	5	10	13	5	10	17	20	49	99
Tithonian	74	394	1268	4	9	20	10	34	66	19	48	100
Toarcian	179	281	593	8	26	60	13	40	110	16	42	85
End-Triassic	3	3	3	7	15	29	2	8	19	12	32	66
End-Permian	11	23	57	51	76	100	8	9	19	21	52	85
Guadalupian	5	40	80	29	82	157	2	7	25	15	34	60
Early-Serpukhovian	7	52	120	3	62	189	17	36	67	24	58	106
End-Devonian	54	167	240	34	192	526	22	87	106	21	52	107
Givetian-Frasnian	194	327	368	27	74	98	6	32	102	26	62	111
Eifelian-Givetian	13	19	44	38	89	103	29	144	156	13	33	66
End-Ordovician	23	37	65	13	29	73	18	48	58	19	46	87
Series 3-Furogian	5	10	10	64	197	387	38	66	73	17	42	85
Stage 3-Stage 4	3	3	8	184	367	435	6	23	82	14	35	74

Table 3: Environmental Variable Bin Summary Table

	n	mean	sd	median	min	max	range	skew	kurtosis	se
Oxygen Isotopes										
2 Myr	16	139.06	198.52	38.5	3	680	677	1.45	1.04	49.63
5 Myr	16	318.25	505.49	109.5	3	1983	1980	2.20	4.47	126.37
10 Myr	16	619.31	1144.17	180.0	3	4603	4600	2.59	6.23	286.04
Carbon Isotopes										
2 Myr	16	73.19	129.63	28.0	3	509	506	2.39	5.08	32.41
5 Myr	16	208.06	409.91	75.0	9	1678	1669	2.84	7.36	102.48
10 Myr	16	417.00	899.64	101.5	13	3696	3683	2.99	8.05	224.91
Strontium Isotopes										
2 Myr	16	13.94	10.95	11.5	2	38	36	0.75	-0.72	2.74
5 Myr	16	47.81	46.34	35.0	7	159	152	1.27	0.42	11.59
10 Myr	16	85.00	88.30	66.5	17	380	363	2.23	4.91	22.07
Sea Level										
2 Myr	16	18.88	4.10	19.5	12	26	14	-0.06	-1.15	1.02
5 Myr	16	46.69	9.64	48.5	32	62	30	-0.11	-1.31	2.41
10 Myr	16	90.56	17.10	93.0	60	111	51	-0.39	-1.35	4.27

Tables 2 and 3 | Number of abiotic variable datapoints per bin and their summary table for the models using three-timer extinction rates.

Number of datapoints for all abiotic variables used in the models using three-timer extinction rates per bin (Table 2). Summary statistics for each bin size for all abiotic variables (Table 3).

instability (based on *s* SL) and extinction enhancement were these data to increase in resolution moving forward.

Model Output and Model Assumption Analysis

Models with Proportion Extinction Rates

Bambach[22] extinction rates were calculated using proportion of genus extinct to total genus diversity within a bin. The bins were on average ~ 3 Myr in length. All the same calculations made for the proxies in the models using three-timer extinction rates were made for the proportion rates. However, where bins start and end are very different between proportion and three-timer extinction rates due to different bin sizes and changes made to the international chronostratigraphic chart between publication dates (an eight year difference).

The models using proportion extinction rates all performed poorly (Bonferroni corrected P-values for 2 Myr-prior: $P = 1$, 5 Myr-prior: $P = 0.42$, and 10 Myr-prior: $P = 0.21$). The 2 Myr-prior model was built with only sixteen observations because there are missing data in some bins due to small sample size.

Models with Three-timer Extinction Rates

All the models using Alroy[6] three-timer extinction rates performed well (Bonferroni corrected P-values for 2 Myr: $P = 0.01$, 5 Myr: $P = 0.003$, 10 Myr: $P = 0.04$, see SI). There were no gross violations of the assumptions of multiple linear regression. However, Cook's Distance revealed the likelihood of outlying data from the end-Permian observation in the 5 Myr-prior model > 2.0 and Stage 3 – Stage 4 observation in the 10 Myr-prior model at ~ 0.9.

Only the 5 Myr-prior model required the addition of an interaction term. The *s* δ^{18}O, *s* δ^{13}C, and *s* SL predictors all initially were significant in the model. After including all interaction effects, only *s* δ^{18}O x *s* SL significantly contributed to the model, indicating the need for its inclusion. The final 5 Myr-prior model resulted in six predictors as opposed to the original five.

Model Assumptions

All datasets were initially assessed for linear relationships using a combination of scatterplots, correlation values, and distribution matrices of kernel density plots (in R code). Six multiple linear regression models were constructed in R using the lm() function. The assumptions of multiple linear regression (i.e., linear relationship between independent and dependent variables and that residual terms are independent, have constant variance, and are normally distributed[32]) were visually assessed using residual vs. fitted value plots, Cook's distance plots, residual vs. predictor plots, QQ-plots, standardized residual vs. observation and leverage vs. observation plots, leverages checked for suspect values, histogram of standardized residuals, autocorrelation of residuals plots, and interaction plots between statistically significant predictors after models were examined (Figs. 4-9). Furthermore, these assumptions were assessed using the Shapiro-Wilk test for normality, Breusch-Pagan test for homogenized residuals variance, Breusch-Godfrey test for serial correlation of order up to 1, variance inflation factor (VIF) analysis for multicollinearity, and Bonferroni outlier test to determine extreme standardized residuals (Table 4).

No gross violations of the assumptions of multiple linear regression were found. However, independence of the residual terms was likely violated because these proxies represent different aspects of the Earth system that influence one other and some isotopic values in these data have been derived from the same sample (e.g., the same bivalve used for $\delta^{18}O$, $\delta^{13}C$, or $^{87}Sr/^{86}Sr$ analyses).

Proportion Extinction Rates

2 Myr-prior model: The assumptions of multiple linear regression were not grossly violated for the 2 Myr-prior model using proportion extinction rates. Of the assessable assumptions, normality of the residuals comes closest to being violated (Fig. 4, K). However, the Shapiro-Wilk test and the normality plot suggest this was a weak violation (P = 0.90, Fig. 4, H). No evidence was found suggesting inconstant variance of the residuals (Fig.4, Table 4). The Capitanian extinction had the highest Cook's distance, < 0.8. The analyses of the standardized residuals and leverages do not suggest any outlying data. VIF analysis resulted in low values, but $\bar{x}$ $\delta^{18}O$ and *m* $^{87}Sr/^{86}Sr$ show slightly higher multicollinearity than the other abiotic variables (Table 4).

5 Myr-prior model: No violations of the assumptions were found in this model (Fig. 5). Again, Cook's distance (> 0.8) for the Capitanian extinction was relatively higher than the other observations, but not alarming.

10 Myr-prior model: No violations of the assumptions were found in the 10 Myr-prior model (Fig. 6). The early Dresbachian may be somewhat of an influential point, but it was calculated at only ~ 0.6, warranting no concern.

Three-timer Extinction Rates

2 Myr-prior model: The 2 Myr-prior model using three-timer extinction rates shows no strong evidence for the violation of the assumptions of multiple linear regression (Fig. 7, Table 4). Cook's distance for the Stage 3 – Stage 4 and Series 3 – Furogian extinction intervals were relatively high, but both being < 0.8 is not concerning. All VIF values calculated for this model are below two (Table 4).

5 Myr-prior model: The 5 Myr-prior model does not seem to be in gross violation of the assumptions of multiple linear regression. There is evidence for the violation of the assumption of normally distributed residual terms (Fig. 8, H and K). However, the Shapiro-Wilk test for normality provided evidence against the violation of this assumption. The Cook's distance (> 2.0) for the end-Permian provides evidence for outlying data (Fig. 8, B). The standardized residual and leverage vs observation plots suggest that there are no datapoints far outside their distributions (Fig. 8, I and J). This model required the inclusion of the interaction term $s\ \delta^{18}O \times s\ SL$, which greatly increased VIF values (Table 4), indicating that this model might be over-complicated.

10 Myr-prior model: There were no gross violations of the assumptions of multiple linear regression found for the 10 Myr-prior model. Stage 3 – Stage 4 is a highly influential point (Cook's distance > 1.0), and both Stage 3 – Stage 4 and Series 3 – Furogian seem to have higher leverage values compared to their distribution (Fig. 4, B and J). However, these results are not extreme enough to require any transformations of the data.

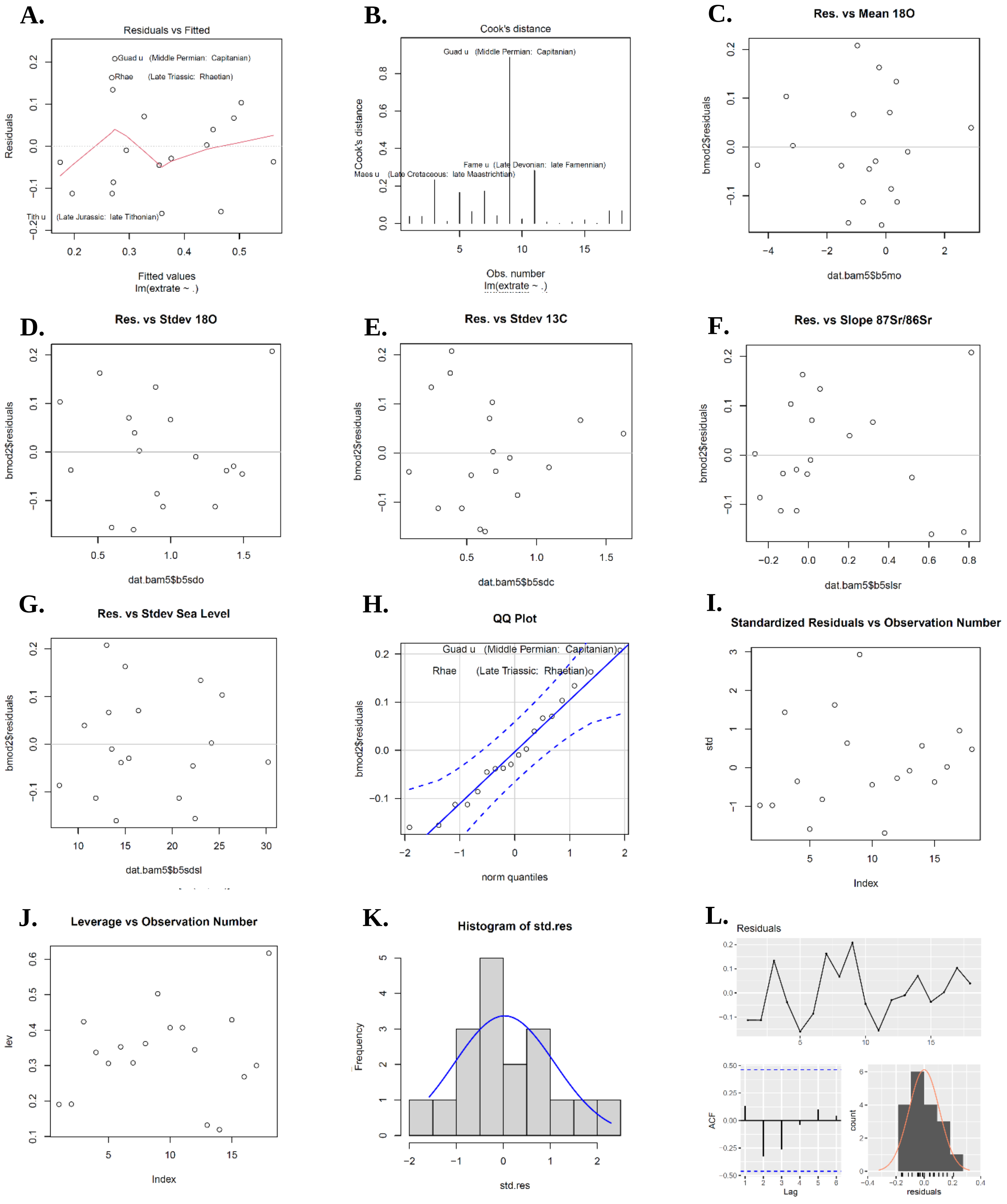


Fig 5 | Proportion extinction rate 5 Myr-prior model plots for model assumption assessment. The assumptions that the residual terms have constant variance and are normally distributed, as well as investigation of possible outliers for multiple linear regression were assessed using the plots above. **A.** residual vs. fitted, **B.** Cook's distance, **C.- G.** residuals vs x̅ δ^{18}O, *s* δ^{18}O, *s* δ^{13}C, *m* ^{87}Sr/^{86}Sr, and *s* SL, **H.** normal probability, **I.** and **J.** standardized residuals and leverage vs observation number, **K.** histogram and kernel density estimation of standardized residuals, **L.** autocorrelation analysis of residuals for one lag plots.

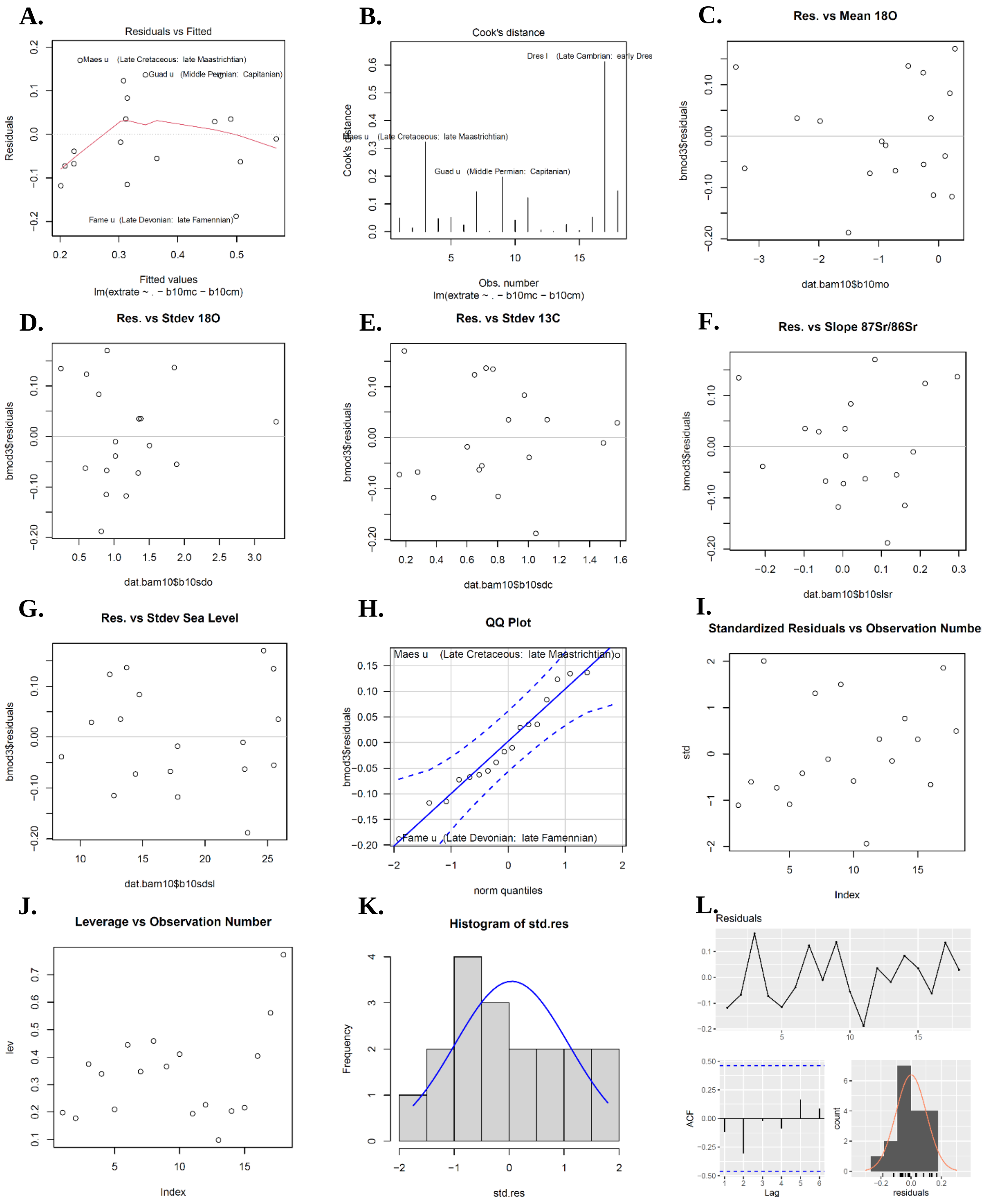


Fig 6 | Proportion extinction rate 10 Myr-prior model plots for model assumption assessment.

The assumptions that the residual terms have constant variance and are normally distributed, as well as investigation of possible outliers for multiple linear regression were assessed using the plots above. **A.** residual vs. fitted, **B.** Cook's distance, **C.- G.** residuals vs $\bar{x}$ $\delta^{18}O$, *s* $\delta^{18}O$, *s* $\delta^{13}C$, *m* $^{87}Sr/^{86}Sr$, and *s* SL, **H.** normal probability, **I.** and **J.** standardized residuals and leverage vs observation number, **K.** histogram and kernel density estimation of standardized residuals, **L.** autocorrelation analysis of residuals for one lag plots.

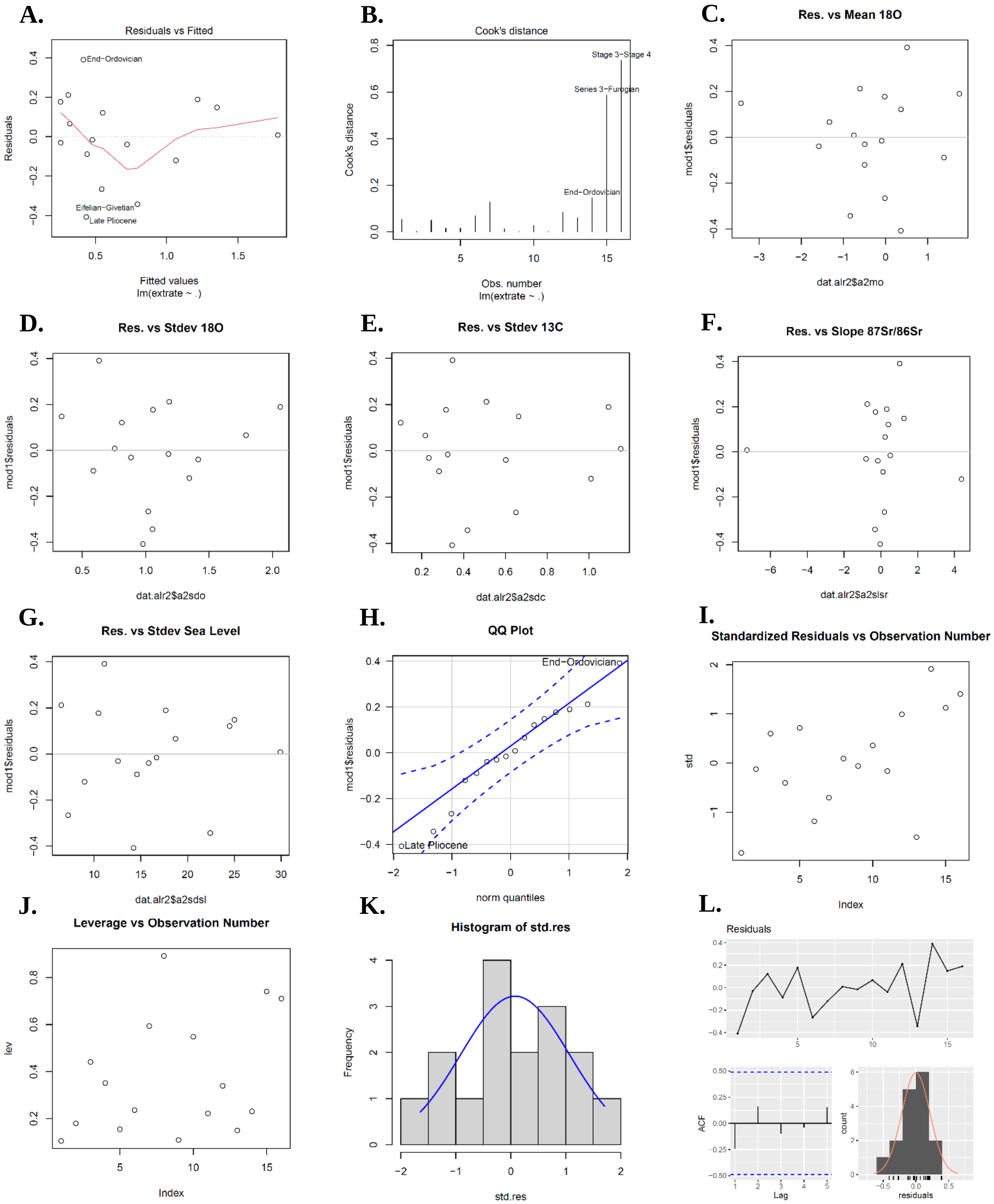


Fig 7 | Three-timer extinction rate 2 Myr-prior model plots for model assumption assessment.

The assumptions that the residual terms have constant variance and are normally distributed, as well as investigation of possible outliers for multiple linear regression were assessed using the plots above. **A.** residual vs. fitted, **B.** Cook's distance, **C.- G.** residuals vs $\bar{x}$ $\delta^{18}O$, *s* $\delta^{18}O$, *s* $\delta^{13}C$, *m* $^{87}Sr/^{86}Sr$, and *s* SL, **H.** normal probability, **I.** and **J.** standardized residuals and leverage vs observation number, **K.** histogram and kernel density estimation of standardized residuals, **L.** autocorrelation analysis of residuals for one lag plots.

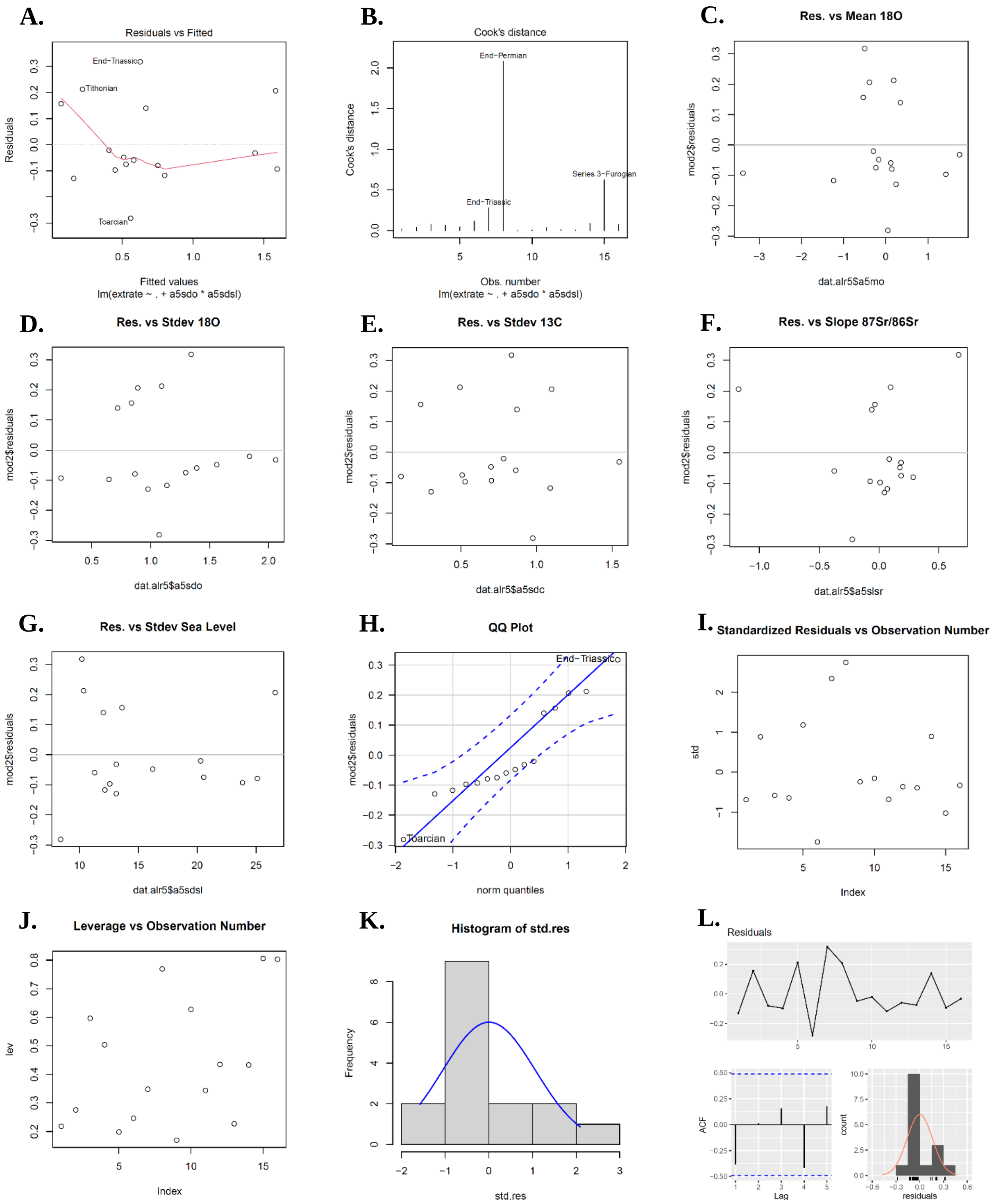


Fig 8 | Three-timer extinction rate 5 Myr-prior model plots for model assumption assessment.

The assumptions that the residual terms have constant variance and are normally distributed, as well as investigation of possible outliers for multiple linear regression were assessed using the plots above. **A.** residual vs. fitted, **B.** Cook's distance, **C.- G.** residuals vs $\bar{x}$ δ^{18}O, *s* δ^{18}O, *s* δ^{13}C, *m* ^{87}Sr/^{86}Sr, and *s* SL, **H.** normal probability, **I.** and **J.** standardized residuals and leverage vs observation number, **K.** histogram and kernel density estimation of standardized residuals, **L.** autocorrelation analysis of residuals for one lag plots.

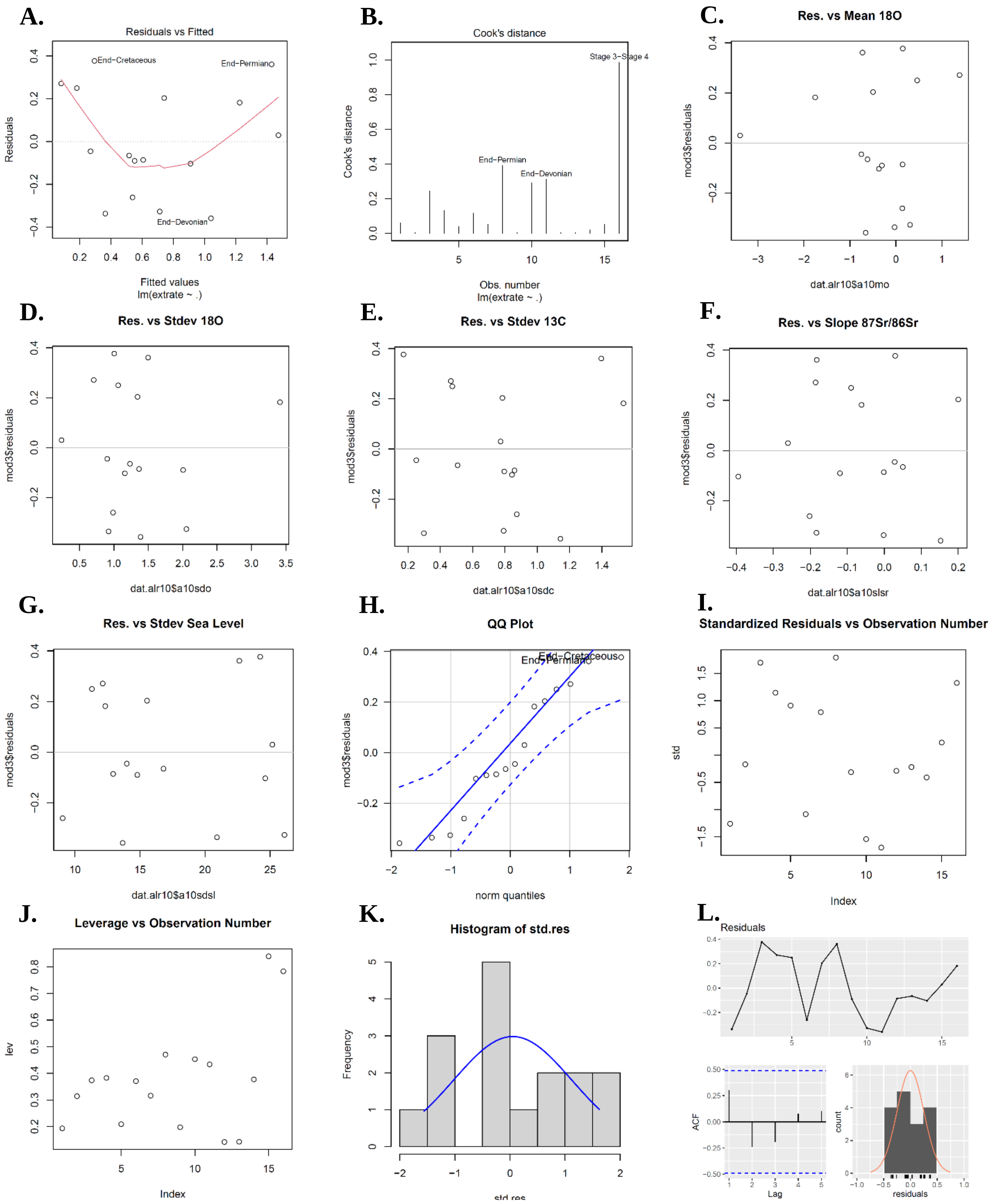
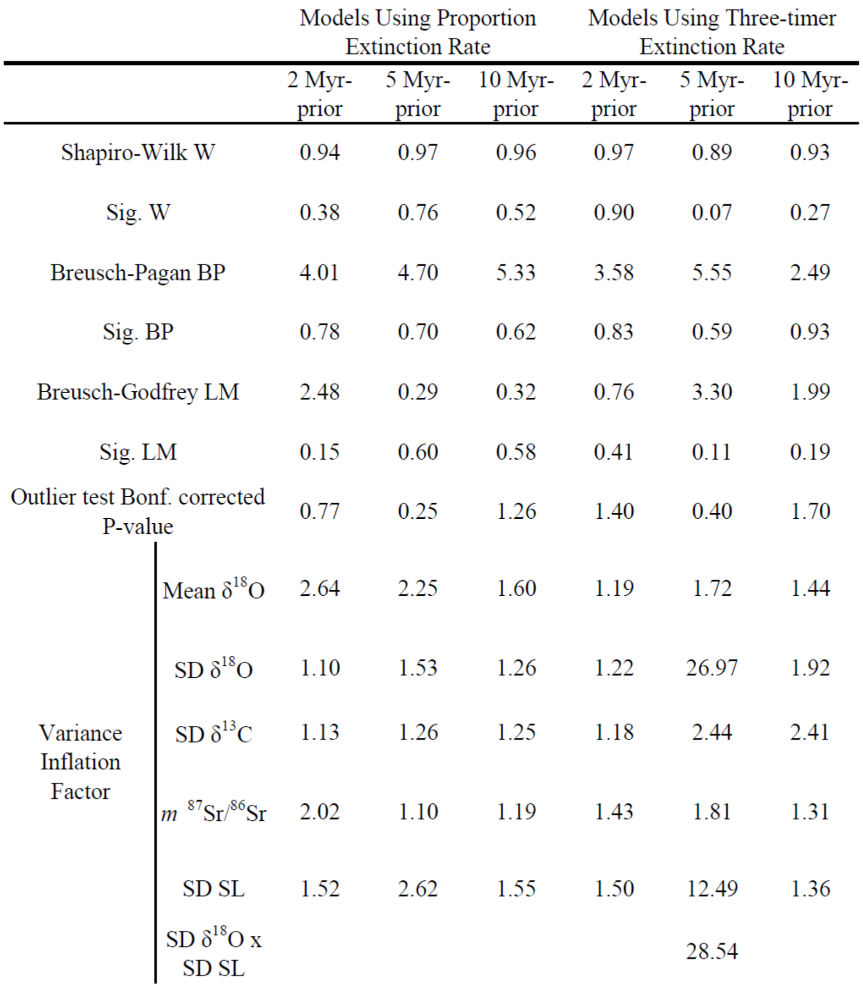


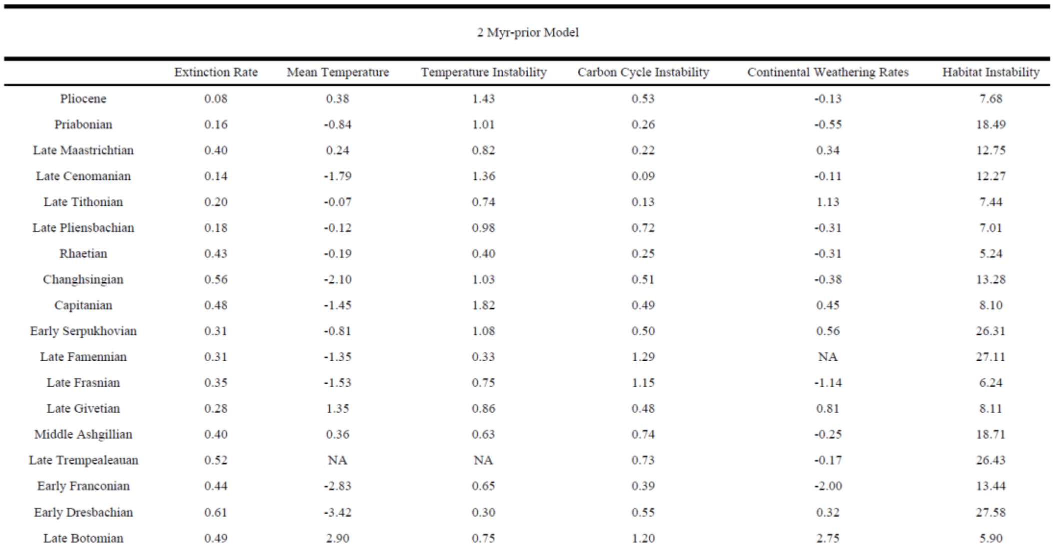
Fig 9 | Three-timer extinction rate 10 Myr-prior model plots for model assumption assessment.

The assumptions that the residual terms have constant variance and are normally distributed, as well as investigation of possible outliers for multiple linear regression were assessed using the plots above. **A.** residual vs. fitted, **B.** Cook's distance, **C.- G.** residuals vs $\bar{x}$ δ^{18}O, *s* δ^{18}O, *s* δ^{13}C, *m* ^{87}Sr/^{86}Sr, and *s* SL, **H.** normal probability, **I.** and **J.** standardized residuals and leverage vs observation number, **K.** histogram and kernel density estimation of standardized residuals, **L.** autocorrelation analysis of residuals for one lag plots.



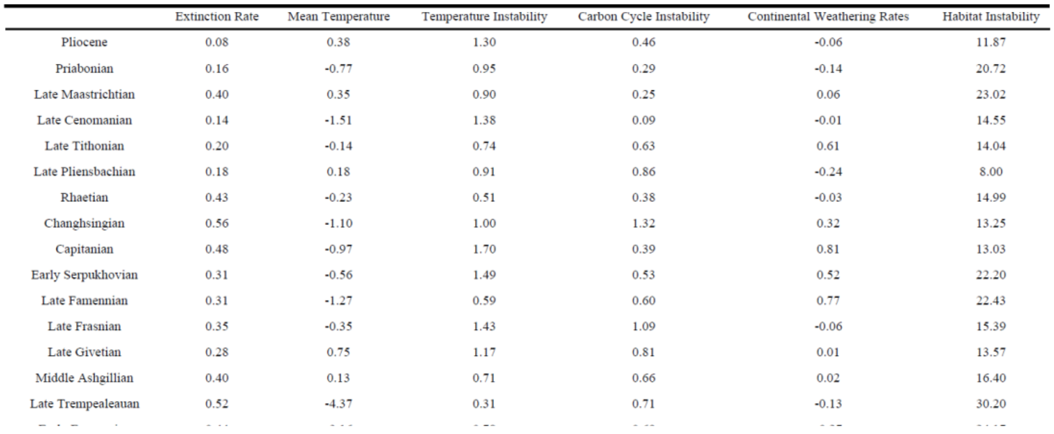
		Models Using Proportion Extinction Rate			Models Using Three-timer Extinction Rate		
		2 Myr-prior	5 Myr-prior	10 Myr-prior	2 Myr-prior	5 Myr-prior	10 Myr-prior
Shapiro-Wilk W		0.94	0.97	0.96	0.97	0.89	0.93
Sig. W		0.38	0.76	0.52	0.90	0.07	0.27
Breusch-Pagan BP		4.01	4.70	5.33	3.58	5.55	2.49
Sig. BP		0.78	0.70	0.62	0.83	0.59	0.93
Breusch-Godfrey LM		2.48	0.29	0.32	0.76	3.30	1.99
Sig. LM		0.15	0.60	0.58	0.41	0.11	0.19
Outlier test Bonf. corrected P-value		0.77	0.25	1.26	1.40	0.40	1.70
Variance Inflation Factor	Mean $\delta^{18}O$	2.64	2.25	1.60	1.19	1.72	1.44
	SD $\delta^{18}O$	1.10	1.53	1.26	1.22	26.97	1.92
	SD $\delta^{13}C$	1.13	1.26	1.25	1.18	2.44	2.41
	m $^{87}Sr/^{86}Sr$	2.02	1.10	1.19	1.43	1.81	1.31
	SD SL	1.52	2.62	1.55	1.50	12.49	1.36
	SD $\delta^{18}O$ x SD SL					28.54	

Table 4 | Statistical tests used to assess for violation of model assumptions. Tests used: Shapiro-Wilk W test for normality of the residuals, Breusch-Pagan BP test for constant variance of the residuals, Breusch-Godfrey LM test to determine degree of residual dependence, Outlier test with Bonferroni corrected P-values of the most extreme standardized residuals (from left to right: Capitanian, Capitanian, late Maastrichtian, end-Ordovician, end-Permian, and end-Permian, and VIF of the model predictors.

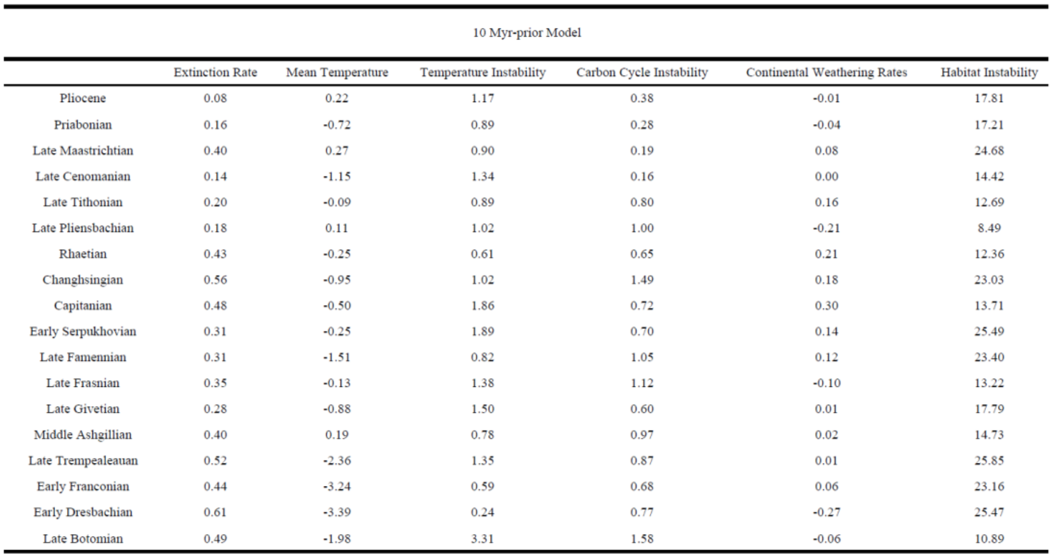


2 Myr-prior Model

	Extinction Rate	Mean Temperature	Temperature Instability	Carbon Cycle Instability	Continental Weathering Rates	Habitat Instability
Pliocene	0.08	0.38	1.43	0.53	-0.13	7.68
Priabonian	0.16	-0.84	1.01	0.26	-0.55	18.49
Late Maastrichtian	0.40	0.24	0.82	0.22	0.34	12.75
Late Cenomanian	0.14	-1.79	1.36	0.09	-0.11	12.27
Late Tithonian	0.20	-0.07	0.74	0.13	1.13	7.44
Late Pliensbachian	0.18	-0.12	0.98	0.72	-0.31	7.01
Rhaetian	0.43	-0.19	0.40	0.25	-0.31	5.24
Changhsingian	0.56	-2.10	1.03	0.51	-0.38	13.28
Capitanian	0.48	-1.45	1.82	0.49	0.45	8.10
Early Serpukhovian	0.31	-0.81	1.08	0.50	0.56	26.31
Late Famennian	0.31	-1.35	0.33	1.29	NA	27.11
Late Frasnian	0.35	-1.53	0.75	1.15	-1.14	6.24
Late Givetian	0.28	1.35	0.86	0.48	0.81	8.11
Middle Ashgillian	0.40	0.36	0.63	0.74	-0.25	18.71
Late Trempealeauan	0.52	NA	NA	0.73	-0.17	26.43
Early Franconian	0.44	-2.83	0.65	0.39	-2.00	13.44
Early Dresbachian	0.61	-3.42	0.30	0.55	0.32	27.58
Late Botomian	0.49	2.90	0.75	1.20	2.75	5.90



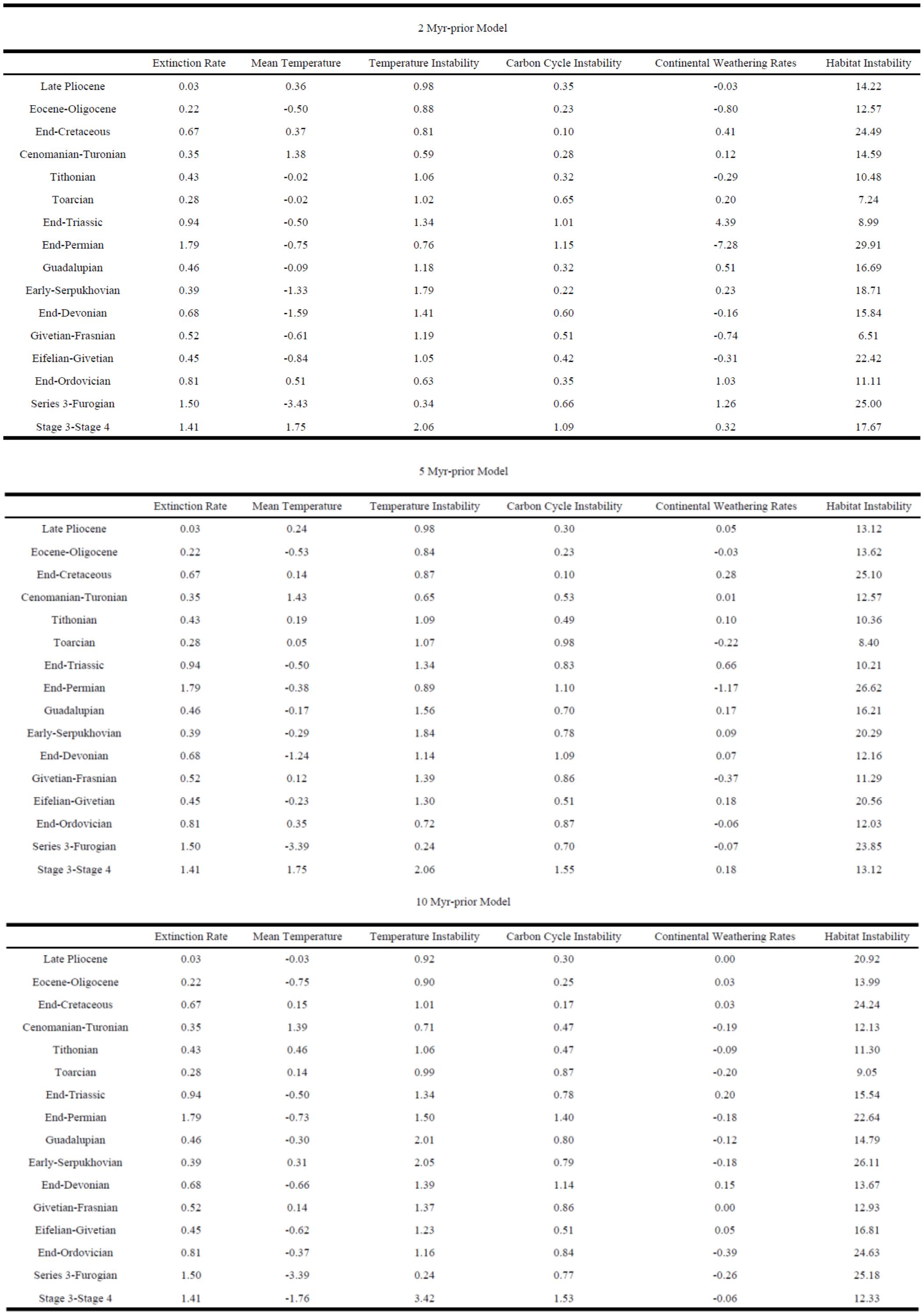
	Extinction Rate	Mean Temperature	Temperature Instability	Carbon Cycle Instability	Continental Weathering Rates	Habitat Instability
Pliocene	0.08	0.38	1.30	0.46	-0.06	11.87
Priabonian	0.16	-0.77	0.95	0.29	-0.14	20.72
Late Maastrichtian	0.40	0.35	0.90	0.25	0.06	23.02
Late Cenomanian	0.14	-1.51	1.38	0.09	-0.01	14.55
Late Tithonian	0.20	-0.14	0.74	0.63	0.61	14.04
Late Pliensbachian	0.18	0.18	0.91	0.86	-0.24	8.00
Rhaetian	0.43	-0.23	0.51	0.38	-0.03	14.99
Changhsingian	0.56	-1.10	1.00	1.32	0.32	13.25
Capitanian	0.48	-0.97	1.70	0.39	0.81	13.03
Early Serpukhovian	0.31	-0.56	1.49	0.53	0.52	22.20
Late Famennian	0.31	-1.27	0.59	0.60	0.77	22.43
Late Frasnian	0.35	-0.35	1.43	1.09	-0.06	15.39
Late Givetian	0.28	0.75	1.17	0.81	0.01	13.57
Middle Ashgillian	0.40	0.13	0.71	0.66	0.02	16.40
Late Trempealeauan	0.52	-4.37	0.31	0.71	-0.13	30.20



10 Myr-prior Model

	Extinction Rate	Mean Temperature	Temperature Instability	Carbon Cycle Instability	Continental Weathering Rates	Habitat Instability
Pliocene	0.08	0.22	1.17	0.38	-0.01	17.81
Priabonian	0.16	-0.72	0.89	0.28	-0.04	17.21
Late Maastrichtian	0.40	0.27	0.90	0.19	0.08	24.68
Late Cenomanian	0.14	-1.15	1.34	0.16	0.00	14.42
Late Tithonian	0.20	-0.09	0.89	0.80	0.16	12.69
Late Pliensbachian	0.18	0.11	1.02	1.00	-0.21	8.49
Rhaetian	0.43	-0.25	0.61	0.65	0.21	12.36
Changhsingian	0.56	-0.95	1.02	1.49	0.18	23.03
Capitanian	0.48	-0.50	1.86	0.72	0.30	13.71
Early Serpukhovian	0.31	-0.25	1.89	0.70	0.14	25.49
Late Famennian	0.31	-1.51	0.82	1.05	0.12	23.40
Late Frasnian	0.35	-0.13	1.38	1.12	-0.10	13.22
Late Givetian	0.28	-0.88	1.50	0.60	0.01	17.79
Middle Ashgillian	0.40	0.19	0.78	0.97	0.02	14.73
Late Trempealeauan	0.52	-2.36	1.35	0.87	0.01	25.85
Early Franconian	0.44	-3.24	0.59	0.68	0.06	23.16
Early Dresbachian	0.61	-3.39	0.24	0.77	-0.27	25.47
Late Botomian	0.49	-1.98	3.31	1.58	-0.06	10.89

Table 5 | Datasets used in model construction testing the relationship between proportion extinction rates and abiotic variables.



2 Myr-prior Model

	Extinction Rate	Mean Temperature	Temperature Instability	Carbon Cycle Instability	Continental Weathering Rates	Habitat Instability
Late Pliocene	0.03	0.36	0.98	0.35	-0.03	14.22
Eocene-Oligocene	0.22	-0.50	0.88	0.23	-0.80	12.57
End-Cretaceous	0.67	0.37	0.81	0.10	0.41	24.49
Cenomanian-Turonian	0.35	1.38	0.59	0.28	0.12	14.59
Tithonian	0.43	-0.02	1.06	0.32	-0.29	10.48
Toarcian	0.28	-0.02	1.02	0.65	0.20	7.24
End-Triassic	0.94	-0.50	1.34	1.01	4.39	8.99
End-Permian	1.79	-0.75	0.76	1.15	-7.28	29.91
Guadalupian	0.46	-0.09	1.18	0.32	0.51	16.69
Early-Serpukhovian	0.39	-1.33	1.79	0.22	0.23	18.71
End-Devonian	0.68	-1.59	1.41	0.60	-0.16	15.84
Givetian-Frasnian	0.52	-0.61	1.19	0.51	-0.74	6.51
Eifelian-Givetian	0.45	-0.84	1.05	0.42	-0.31	22.42
End-Ordovician	0.81	0.51	0.63	0.35	1.03	11.11
Series 3-Furogian	1.50	-3.43	0.34	0.66	1.26	25.00
Stage 3-Stage 4	1.41	1.75	2.06	1.09	0.32	17.67

5 Myr-prior Model

	Extinction Rate	Mean Temperature	Temperature Instability	Carbon Cycle Instability	Continental Weathering Rates	Habitat Instability
Late Pliocene	0.03	0.24	0.98	0.30	0.05	13.12
Eocene-Oligocene	0.22	-0.53	0.84	0.23	-0.03	13.62
End-Cretaceous	0.67	0.14	0.87	0.10	0.28	25.10
Cenomanian-Turonian	0.35	1.43	0.65	0.53	0.01	12.57
Tithonian	0.43	0.19	1.09	0.49	0.10	10.36
Toarcian	0.28	0.05	1.07	0.98	-0.22	8.40
End-Triassic	0.94	-0.50	1.34	0.83	0.66	10.21
End-Permian	1.79	-0.38	0.89	1.10	-1.17	26.62
Guadalupian	0.46	-0.17	1.56	0.70	0.17	16.21
Early-Serpukhovian	0.39	-0.29	1.84	0.78	0.09	20.29
End-Devonian	0.68	-1.24	1.14	1.09	0.07	12.16
Givetian-Frasnian	0.52	0.12	1.39	0.86	-0.37	11.29
Eifelian-Givetian	0.45	-0.23	1.30	0.51	0.18	20.56
End-Ordovician	0.81	0.35	0.72	0.87	-0.06	12.03
Series 3-Furogian	1.50	-3.39	0.24	0.70	-0.07	23.85
Stage 3-Stage 4	1.41	1.75	2.06	1.55	0.18	13.12

10 Myr-prior Model

	Extinction Rate	Mean Temperature	Temperature Instability	Carbon Cycle Instability	Continental Weathering Rates	Habitat Instability
Late Pliocene	0.03	-0.03	0.92	0.30	0.00	20.92
Eocene-Oligocene	0.22	-0.75	0.90	0.25	0.03	13.99
End-Cretaceous	0.67	0.15	1.01	0.17	0.03	24.24
Cenomanian-Turonian	0.35	1.39	0.71	0.47	-0.19	12.13
Tithonian	0.43	0.46	1.06	0.47	-0.09	11.30
Toarcian	0.28	0.14	0.99	0.87	-0.20	9.05
End-Triassic	0.94	-0.50	1.34	0.78	0.20	15.54
End-Permian	1.79	-0.73	1.50	1.40	-0.18	22.64
Guadalupian	0.46	-0.30	2.01	0.80	-0.12	14.79
Early-Serpukhovian	0.39	0.31	2.05	0.79	-0.18	26.11
End-Devonian	0.68	-0.66	1.39	1.14	0.15	13.67
Givetian-Frasnian	0.52	0.14	1.37	0.86	0.00	12.93
Eifelian-Givetian	0.45	-0.62	1.23	0.51	0.05	16.81
End-Ordovician	0.81	-0.37	1.16	0.84	-0.39	24.63
Series 3-Furogian	1.50	-3.39	0.24	0.77	-0.26	25.18
Stage 3-Stage 4	1.41	-1.76	3.42	1.53	-0.06	12.33

Table 6 | Datasets used in model construction testing the relationship between three-timer extinction rates and abiotic variables.

References

1. Prokoph, A., Shields, G. A. & Veizer, J. Compilation and time-series analysis of a marine carbonate δ18O, δ13C, 87Sr/86Sr and δ34S database through Earth history. *Earth-Science Rev.* **87**, 113–133 (2008).
2. Veizer, J. & Prokoph, A. Temperatures and oxygen isotopic composition of Phanerozoic oceans. *Earth-Science Rev.* **146**, 92–104 (2015).
3. Bachan, A. *et al.* A model for the decrease in amplitude of carbon isotope excursions across the phanerozoic. *Am. J. Sci.* **317**, 641–676 (2017).
4. Boulila, S., Laskar, J., Haq, B. U., Galbrun, B. & Hara, N. Long-term cyclicities in Phanerozoic sea-level sedimentary record and their potential drivers. *Glob. Planet. Change* **165**, 128–136 (2018).
5. Zaky, A. H. *et al.* Strontium isotope geochemistry of modern and ancient archives: Tracer of secular change in ocean chemistry. *Can. J. Earth Sci.* **56**, 245–264 (2019).
6. Alroy, J. Accurate and precise estimates of origination and extinction rates. *Paleobiology* **40**, 374–397 (2014).
7. Finnegan, S. *et al.* Paleontological baselines for evaluating extinction risk in the modern oceans. *Science (80-.).* **348**, 567–570 (2015).
8. Barnosky, A. D. *et al.* Has the Earth’s sixth mass extinction already arrived? *Nature* **471**, 51–57 (2011).
9. Hull, P. M., Darroch, S. A. F. & Erwin, D. H. Rarity in mass extinctions and the future of ecosystems. *Nature* **528**, 345–351 (2015).
10. Dietl, G. P. & Flessa, K. W. Conservation paleobiology: Putting the dead to work. *Trends Ecol. Evol.* **26**, 30–37 (2011).

11. Sepkoski, J. J. A Kinetic Model of Phanerozoic Taxonomic Diversity . III . Post-Paleozoic Families and Mass Extinctions. **10**, 246–267 (1984).

12. Raup, D. M. & Sepkoski, J. J. Periodic extinction of families and genera. *Science (80-.).* **231**, 833–836 (1986).

13. Alroy, J. Dynamics of origination and extinction in the marine fossil record. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11536–11542 (2008).

14. MacArthur, R. H. & Wilson, E. O. An Equilibrium Theory of Insular Zoogeography. *Evolution (N. Y).* **17**, 373 (1963).

15. Smth, J. M. The causes of extinction. *Philos. Trans. R. Soc. B Biol. Sci.* **325**, 241–252 (1989).

16. Benton, M. J. The Red Queen and the Court Jester: Species diversity and the role of biotic and abiotic factors through time. *Science (80-.).* **323**, 728–732 (2009).

17. Sepkoski, J. J. A kinetic model of Phanerozoic taxonomic diversity II . Early Phanerozoic families and multiple equilibria. **5**, 222–251 (1979).

18. Foote, M., Cooper, R. A., Crampton, J. S. & Sadler, P. M. Diversity-dependent evolutionary rates in early palaeozoic zooplankton. *Proc. R. Soc. B Biol. Sci.* **285**, 11–14 (2018).

19. Peters, S. E. Environmental determinants of extinction selectivity in the fossil record. *Nature* **454**, 626–629 (2008).

20. Alroy, J. Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology* **53**, 1211–1235 (2010).

21. Cárdenas, A. L. & Harries, P. J. Effect of nutrient availability on marine origination rates throughout the Phanerozoic eon. *Nat. Geosci.* **3**, 430–434 (2010).

22. Bambach, R. K. Phanerozoic Biodiversity Mass Extinctions. *Annu. Rev. Earth Planet. Sci.* **34**, 127–155 (2006).

23. Sepkoski, J. J. A Compendium of Fossil Marine Animal Genera. *Bull. Am. Paleontol.* **363**, 1–560 (2002).

24. Alroy, J. *et al.* Phanerozoic trends in the global diversity of marine invertebrates. *Science (80-.).* **321**, 97–100 (2008).

25. Kump, L. R., Brantley, S. L. & Arthur, M. A. Chemical Weathering, Atmospheric CO2, and Climate. *Annu. Rev. Earth Planet. Sci.* **28**, 611–667 (2000).

26. Airoldi, L., Balata, D. & Beck, M. W. The Gray Zone: Relationships between habitat loss and marine diversity and their applications in conservation. *J. Exp. Mar. Bio. Ecol.* **366**, 8–15 (2008).

27. Jorda, G. *et al.* Ocean warming compresses the three-dimensional habitat of marine life. *Nat. Ecol. Evol.* **4**, 109–114 (2020).

28. R Core Team. R: A language and environment for statistical computing. (2021).

29. Veizer, J., Godderis, Y. & François, L. M. Evidence for decoupling of atmospheric CO2 and global climate during the Phanerozoic eon. *Nature* **408**, 698–701 (2000).

30. Ryb, U. & Eiler, J. M. Oxygen isotope composition of the Phanerozoic ocean and a possible solution to the dolomite problem. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6602–6607 (2018).

31. Hearing, T. W. *et al.* An early Cambrian greenhouse climate. *Sci. Adv.* **4**, 1–12 (2018).

32. Christensen, R. *Analysis of Variance, Design, and Regression: Linear Modeling for Unbalanced Data*. (CRC Press, Taylor & Francis Group, 2016).

33. Allen, D. M. The Relationship Between Variable Selection and Data Agumentation

and a Method for Prediction. *Technometrics* **16**, 125–127 (1974).

34. Erwin, D. H., Bowring, S. A. & Yugan, J. End-Permian mass extinctions: A review. *Spec. Pap. Geol. Soc. Am.* **356**, 363–383 (2002).
35. Stenseth, N. C. *et al.* Ecological effects of climate fluctuations. *Science (80-.).* **297**, 1292–1296 (2002).
36. Allstadt, A. J., Liebhold, A. M., Johnson, D. M., Davis, R. E. & Haynes, K. J. Temporal variation in the synchrony of weather and its consequences for spatiotemporal population dynamics. *Ecology* **96**, 2935–2946 (2015).
37. Hansen, B. B., Grøtan, V., Herfindal, I. & Lee, A. M. The Moran effect revisited: spatial population synchrony under global warming. *Ecography (Cop.).* **43**, 1591–1602 (2020).
38. Moran, P. A. P. The Statistical Analysis of the Canadian Lynx Cycle II. Synchronization and Meteorology. *Aust. J. Zool.* **1**, 291–298 (1953).
39. Spencer, P. D. Density-independent and density-dependent factors affecting temporal changes in spatial distributions of eastern Bering Sea flatfish. *Fish. Oceanogr.* **17**, 396–410 (2008).
40. Ruddiman, W. F. The Anthropogenic Greenhouse Era. *Clim. Change* 261–293 (2003).
41. Guthrie, R. D. New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature* **441**, 207–209 (2006).
42. Ellis, E. C. Ecology in an anthropogenic biosphere. *Ecol. Monogr.* **85**, 287–331 (2015).
43. Lanoë, F. B., Reuther, J. D., Holmes, C. E. & Potter, B. A. Small mammals and paleovironmental context of the terminal pleistocene and early holocene human

occupation of central Alaska. *Geoarchaeology* **35**, 164–176 (2020).

44. Pimm, S. L. *et al.* The biodiversity of species and their rates of extinction, distribution, and protection. *Science (80-.).* **344**, (2014).
45. Tittensor, D. P. *et al.* Biodiversity Targets. *Science (80-.).* **346**, 241–243 (2014).
46. IPBES. *IPBES Global Assessment Summary for Policymakers*. (2019).
47. Sepkoski, J. J. & Kendrick, D. C. Numerical experiments with model monophyletic and paraphyletic taxa. *Paleobiology* **19**, 168–184 (1993).
48. Foote, M., Ritterbush, K. A. & Miller, A. I. Geographic ranges of genera and their constituent species: Structure, evolutionary dynamics, and extinction resistance. *Paleobiology* **42**, 269–288 (2016).
49. Hendricks, J. R., Saupe, E. E., Myers, C. E., Hermsen, E. J. & Allmon, W. D. The Generification of the Fossil Record. *Paleobiology* **40**, 511–528 (2014).
50. Rothman, D. H. Characteristic disruptions of an excitable carbon cycle. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 14813–14822 (2019).
51. Arnscheidt, C. & Rothman, D. Asymmetric extreme events of Earth’s climate and carbon cycle during the last 66 million years. *Bull. Am. Phys. Soc.* (2021).
52. Saltzman, M. R. & Thomas, E. Carbon isotope stratigraphy. *Geol. Time Scale 2012* 207–232 (2012) doi:10.1016/B978-0-444-59425-9.00011-1.
53. Veizer, J. *et al.* 87Sr/86Sr, δ13C and δ18O evolution of Phanerozoic seawater. *Chem. Geol.* **161**, 59–88 (1999).
54. Veizer, J. Strontium Isotopes in Seawater Through Time Strontium Isotope Systematics. *Ann. Rev. Earth Planet. Sci* **17**, 14167 (1989).
55. Sepkoski, J. J. *Phanerozoic overview of mass extinction*. *Patterns and Processes in the*

History of Life (Springer-Verlag, 1986).

56. Alroy, J. Fair Sampling of Taxonomic Richness and Unbiased Estimation of Origination and Extinction Rates. *Paleontol. Soc. Pap.* **16**, 55–80 (2010).
57. Wold, S. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics* **20**, 397–405 (1978).
58. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109–130 (2001).
59. Clark, M. & Cramer, R. D. The Probability of Chance Correlation Using Partial Least Squares (PLS). *Quant. Struct. Relationships* **12**, 137–145 (1993).
60. Cramer, R. D., Patterson, D. E. & Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *Handb. Chemoinformatics* **110**, 5959–5967 (1988).
61. Tropsha, A. & Cho, S. J. Cross-Validated R2 Guided Region Selection for CoMFA Studies. *Perspect. Drug Discov. Des.* **12**, 57–69 (1998).
62. Golub, G. H., Heath, M. & Wahba, G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**, 215–223 (1979).
63. Walker, L. J., Wilkinson, B. H. & Ivany, L. C. Continental drift and phanerozoic carbonate accumulation in shallow-shelf and deep-marine settings. *J. Geol.* **110**, 75–87 (2002).
64. Foote, M. Environmental controls on geographic range size in marine animal genera. *Paleobiology* **40**, 440–458 (2014).
65. Kocsis, Á. T. & Scotese, C. R. Mapping paleocoastlines and continental flooding during the Phanerozoic. *Earth-Science Rev.* **213**, (2021).

Statistical Software Code

All analyses were conducted within the R Programming environment[28].

```
library(ggplot2)
library(GGally)
library(psych)
library(car)
library(MASS)
library(lmtest)
library(MPV) #PRESS
library(relaimpo)
library(effects)
library(forecast)
library(broom)
library(dplyr)
library(plyr)
library(reshape2)
library(fANCOVA)
library(DataCombine)
library(extrafont)

setwd("F:/Projects/Masters/Boundary Conditions/Analyses/Master Data File")

dat<-read.csv("mstrdf.csv", sep=",",check.names=FALSE,header=TRUE)

#subsetting data for analysis
dat0.1<-dat[,c(3,6:10)]
dat0.2<-dat[,c(3,6:10,33:34)]

#original data
dat1<-dat0.1[-c(100355,106025:106590),]
rownames(dat1)<-NULL
dat1$stron<-dat1$stron*10^4

#computation time ~5 minutes

  #for detrending data
  dat0.1<-dat[,c(3,6:10)]
  dat0<-dat0.1[-c(100355,106025:106590),]
  rownames(dat0)<-NULL
  dat0$stron<-dat0$stron*10^4

  #detrending linear regression and lowess regression

  #resorting dox
```

```
  sort.dox<-(dat0[complete.cases(dat0$dox), ])[,1:2]
  sort.dox<-sort.dox[order(sort.dox$orgage),]
  #resorting dcarb
  sort.dcarb<-(dat0[complete.cases(dat0$dcarb), ])[,c(1,3)]
  sort.dcarb<-sort.dcarb[order(sort.dcarb$orgage),]
  #resorting strontium
  sort.stron<-(dat0[complete.cases(dat0$stron), ])[,c(1,4)]
  sort.stron<-sort.stron[order(sort.stron$orgage),]
  #resorting sea level
  sort.sl<-(dat0[complete.cases(dat0$sl), ])[,c(1,5)]
  sort.sl<-sort.sl[order(sort.sl$orgage),]
  #resorting carbonate
  sort.patc<-(dat0[complete.cases(dat0$patc), ])[,c(1,6)]
  sort.patc<-sort.patc[order(sort.patc$orgage),]

  #loess regression detrending with autospan and symmetric family discounting unusual
points


    #oxygen
    doxloe<-loess.as(sort.dox$orgage,sort.dox$dox,degree=1,criterion=c("aicc","gcv")[2],
                 family=c("gaussian","symmetric")[2],user.span=NULL,plot=FALSE)
    doxloe$pars$span

    #carbon
    dcarbloe<-
loess.as(sort.dcarb$orgage,sort.dcarb$dcarb,degree=1,criterion=c("aicc","gcv")[2],
                   family=c("gaussian","symmetric")[2],user.span=NULL,plot=FALSE)
    dcarbloe$pars$span

    #strontium
    stronloe<-
loess.as(sort.stron$orgage,sort.stron$stron,degree=1,criterion=c("aicc","gcv")[2],
                   family=c("gaussian","symmetric")[2],user.span=NULL,plot=FALSE)
    stronloe$pars$span

    #sea level
    slloe<-loess.as(sort.sl$orgage,sort.sl$sl,degree=1,criterion=c("aicc","gcv")[2],
                family=c("gaussian","symmetric")[2],user.span=NULL,plot=FALSE)
    slloe$pars$span

    #ccs
    patcloe<-loess.as(sort.patc$orgage,sort.patc$patc,degree=1,criterion=c("aicc","gcv")[2],
                  family=c("gaussian","symmetric")[2],user.span=NULL,plot=FALSE)
    patcloe$pars$span
```

```
    #dox detrended
    dox<-as.data.frame(cbind(sort.dox$orgage,doxloe$residuals))
    colnames(dox)<-c("orgage","dox")
    #dcarb detrended
    dcarb<-as.data.frame(cbind(sort.dcarb$orgage,dcarbloe$residuals))
    colnames(dcarb)<-c("orgage","dcarb")
    #ston detrended
    stron<-as.data.frame(cbind(sort.stron$orgage,stronloe$residuals))
    colnames(stron)<-c("orgage","stron")
    #sl detrended
    sl<-as.data.frame(cbind(sort.sl$orgage,slloe$residuals))
    colnames(sl)<-c("orgage","sl")
    #patc detrended
    patc<-as.data.frame(cbind(sort.patc$orgage,patcloe$residuals))
    colnames(patc)<-c("orgage","patc")


  #removing data outside 3sd (99.73% CI)
  #data[!(abs(d3sd$dox - mean(d3sd$dox,na.rm=TRUE))/sd(d3sd$dox.dox,na.rm=TRUE))>
3,]

    dox.sd3<-dox[!abs(scale(dox$dox)) > 3,] #data-(mean.var1/sd.var1)>3sd
  dox.sd3 <- dox.sd3[order(dox.sd3$orgage),]
  rownames(dox.sd3)<-NULL

  dcarb.sd3<-dcarb[!abs(scale(dcarb$dcarb)) > 3,]
  dcarb.sd3 <- dcarb.sd3[order(dcarb.sd3$orgage),]
  rownames(dcarb.sd3)<-NULL

  dcarb.sd3.no<-dat0.1[,c(1,3)][!abs(scale(dat0.1$dcarb)) > 3,]
  dcarb.sd3.no <- dcarb.sd3[order(dcarb.sd3.no$orgage),]
  rownames(dcarb.sd3.no)<-NULL

  stron.sd3<-stron[!abs(scale(stron$stron)) > 3,]
  stron.sd3 <- stron.sd3[order(stron.sd3$orgage),]
  rownames(stron.sd3)<-NULL

  sl.sd3<-sl[!abs(scale(sl$sl)) > 3,]
  sl.sd3 <- sl.sd3[order(sl.sd3$orgage),]
  rownames(sl.sd3)<-NULL

  patc.sd3<-patc[!abs(scale(patc$patc)) > 3,]
  patc.sd3 <- patc.sd3[order(patc.sd3$orgage),]
  rownames(patc.sd3)<-NULL
```

```
  #new dataframe of detrended variables and! outliers removed
    d<-data.frame(orgage=c(dox.sd3[,1],dcarb.sd3[,1],stron.sd3[,1],sl.sd3[,1],patc.sd3[,1]))
    dcom1<-rbind.fill(data.frame(dox.sd3[,2]),data.frame(dcarb.sd3[,2]))
    dcom2<-rbind.fill(dcom1,data.frame(stron.sd3[,2]))
    dcom3<-rbind.fill(dcom2,data.frame(sl.sd3[,2]))
    dcom4<-rbind.fill(dcom3,data.frame(patc.sd3[,2]))

    dat1<-cbind(d,dcom4)
    colnames(dat1)<-c("orgage","dox","dcarb","stron","sl","patc")

#before after detrending plots +autocorrelation
  #oldschool plots, before detrending

    plot(dat0$orgage,dat0$dox)#oxygen isotopes
  plot(dat0$orgage,dat0$dcarb)#carbon isotopes
  plot(dat0.2$orgage,dat0.2$stron)#strontium isotope ratios
  #plot(dat0[-100355,]$orgage,dat0[-100355,]$stron)#strontium isotope ratios without
"extreme" value
  plot(dat0$orgage,dat0$sl)#sea level
  plot(dat0$orgage,dat0$patc)#proportion carbonate

  #detrended plots

    font_import()
    loadfonts(device = "win")
    #pdf(file = "F:/Projects/Masters/Boundary Conditions/Analyses/Results
Plots/Scatterplots/proxies.pdf",width = 11,height = 11)
    windowsFonts(A = windowsFont("Times New Roman"))
    par(mfrow=c(2,2))
    plot(dat1$orgage,dat1$dox,xlim = rev(range(dat1$orgage)),
           xlab="Time (Ma)",ylab="Oxygen Isotopes",col="black",bg="brown",pch=21,
        cex.lab=1.5,cex.axis=1.5,cex.main=1.5,cex.sub=1.5)#oxygen isotopes
    title(main="A.",adj=0,family="serif",font=2)
    plot(dat1$orgage,dat1$dcarb,xlim = rev(range(dat1$orgage)),
        xlab="Time (Ma)",ylab="Carbon Isotopes",col="black",bg="dark green",pch=21,
        cex.lab=1.5,cex.axis=1.5,cex.main=1.5,cex.sub=1.5)#carbon isotopes
    title(main="B.",adj=0,family="serif",font=2)
    plot(dat1$orgage,dat1$stron,xlim = rev(range(dat1$orgage)),
        xlab="Time (Ma)",ylab="Strontium Isotope Ratios",col="black",bg="purple",pch=21,
        cex.lab=1.5,cex.axis=1.5,cex.main=1.5,cex.sub=1.5)#strontium isotope ratios
    title(main="C.",adj=0,family="serif",font=2)
    plot(dat1$orgage,dat1$sl,xlim = rev(range(dat1$orgage)),
        xlab="Time (Ma)",ylab="Sea Level",col="black",bg="dark blue",pch=21,
        cex.lab=1.5,cex.axis=1.5,cex.main=1.5,cex.sub=1.5)#sea level
    title(main="D.",adj=0,family="serif",font=2)
    plot(dat1$orgage,dat1$patc)#proportion carbonate
```

```
  par(mfrow=c(1,1))

#all data counts

  aods<-describe(dat0.1[,1:5],na.rm=TRUE)
  adds<-describe(dat1[,1:5],na.rm=TRUE)
  ads<-data.frame(aods$n,adds$n,row.names = rownames(adds))
  ads$removed <- (ads$aods.n-ads$adds.n)
  tot<-cbind(sum(ads[2:5,1]),sum(ads[2:5,2]),sum(ads[2:5,3]))
  colnames(tot)<-colnames(ads)
  ads<-rbind(ads,tot)
  row.names(ads)[6]<-"tot"

#binned dataframe construction

  #functions
  #Bambach ox iso means
  var1<-function(var,age_col,age_end,age_start){
    m<-rbind(mean(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
             mean(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
             mean(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
             mean(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
             mean(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
             mean(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
             mean(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
             mean(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
             mean(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
             mean(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
             mean(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
             mean(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
             mean(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
             mean(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
             mean(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
             mean(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE),
             mean(var[age_col>=age_end[17] & age_col<=age_start[17]],na.rm=TRUE),
             mean(var[age_col>=age_end[18] & age_col<=age_start[18]],na.rm=TRUE))
    return(m)
  }

  #Bambach ox iso sd
  var2<-function(var,age_col,age_end,age_start){
    s<-rbind(sd(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
             sd(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
             sd(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
             sd(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
```

```
        sd(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
        sd(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
        sd(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
        sd(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
        sd(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
        sd(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
        sd(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
        sd(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
        sd(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
        sd(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
        sd(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
        sd(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE),
        sd(var[age_col>=age_end[17] & age_col<=age_start[17]],na.rm=TRUE),
        sd(var[age_col>=age_end[18] & age_col<=age_start[18]],na.rm=TRUE))
  return(s)
}

#Bambach carb iso sd
var3<-function(var,age_col,age_end,age_start){
  s<-rbind(sd(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
        sd(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
        sd(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
        sd(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
        sd(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
        sd(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
        sd(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
        sd(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
        sd(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
        sd(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
        sd(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
        sd(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
        sd(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
        sd(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
        sd(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
        sd(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE),
        sd(var[age_col>=age_end[17] & age_col<=age_start[17]],na.rm=TRUE),
        sd(var[age_col>=age_end[18] & age_col<=age_start[18]],na.rm=TRUE))
  return(s)
}

#Bambach stron iso sd
var4<-function(var,age_col,age_end,age_start){
  {lm1sr<-lm(var~age_col,subset=(age_col>=age_end[1] & age_col<=age_start[1]))
  lm2sr<-lm(var~age_col,subset=(age_col>=age_end[2] & age_col<=age_start[2]))
  lm3sr<-lm(var~age_col,subset=(age_col>=age_end[3] & age_col<=age_start[3]))
  lm4sr<-lm(var~age_col,subset=(age_col>=age_end[4] & age_col<=age_start[4]))
```

```
  lm5sr<-lm(var~age_col,subset=(age_col>=age_end[5] & age_col<=age_start[5]))
  lm6sr<-lm(var~age_col,subset=(age_col>=age_end[6] & age_col<=age_start[6]))
  lm7sr<-lm(var~age_col,subset=(age_col>=age_end[7] & age_col<=age_start[7]))
  lm8sr<-lm(var~age_col,subset=(age_col>=age_end[8] & age_col<=age_start[8]))
  lm9sr<-lm(var~age_col,subset=(age_col>=age_end[9] & age_col<=age_start[9]))
  lm10sr<-lm(var~age_col,subset=(age_col>=age_end[10] & age_col<=age_start[10]))
  lm11sr<-lm(var~age_col,subset=(age_col>=age_end[11] & age_col<=age_start[11]))
  lm12sr<-lm(var~age_col,subset=(age_col>=age_end[12] & age_col<=age_start[12]))
  lm13sr<-lm(var~age_col,subset=(age_col>=age_end[13] & age_col<=age_start[13]))
  lm14sr<-lm(var~age_col,subset=(age_col>=age_end[14] & age_col<=age_start[14]))
  lm15sr<-lm(var~age_col,subset=(age_col>=age_end[15] & age_col<=age_start[15]))
  lm16sr<-lm(var~age_col,subset=(age_col>=age_end[16] & age_col<=age_start[16]))
  lm17sr<-lm(var~age_col,subset=(age_col>=age_end[17] & age_col<=age_start[17]))
  lm18sr<-lm(var~age_col,subset=(age_col>=age_end[18] & age_col<=age_start[18]))}
  { return(rbind((-1*lm1sr$coefficients[2]),(-1*lm2sr$coefficients[2]),(-
1*lm3sr$coefficients[2]),(-1*lm4sr$coefficients[2]),
         (-1*lm5sr$coefficients[2]),(-1*lm6sr$coefficients[2]),(-
1*lm7sr$coefficients[2]),(-1*lm8sr$coefficients[2]),
         (-1*lm9sr$coefficients[2]),(-1*lm10sr$coefficients[2]),(-
1*lm11sr$coefficients[2]),(-1*lm12sr$coefficients[2]),
         (-1*lm13sr$coefficients[2]),(-1*lm14sr$coefficients[2]),(-
1*lm15sr$coefficients[2]),(-1*lm16sr$coefficients[2]),
         (-1*lm17sr$coefficients[2]),(-1*lm18sr$coefficients[2])))}
 }

 #Bambach sl sd
 var5<-function(var,age_col,age_end,age_start){
  s<-rbind(sd(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
      sd(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
      sd(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
      sd(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
      sd(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
      sd(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
      sd(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
      sd(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
      sd(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
      sd(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
      sd(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
      sd(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
      sd(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
      sd(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
      sd(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
      sd(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE),
      sd(var[age_col>=age_end[17] & age_col<=age_start[17]],na.rm=TRUE),
      sd(var[age_col>=age_end[18] & age_col<=age_start[18]],na.rm=TRUE))
  return(s)
```

```
}

#Bambach carbshelf means
var6<-function(var,age_col,age_end,age_start){
  m<-rbind(mean(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
           mean(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
           mean(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
           mean(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
           mean(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
           mean(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
           mean(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
           mean(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
           mean(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
           mean(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
           mean(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
           mean(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
           mean(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
           mean(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
           mean(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
           mean(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE),
           mean(var[age_col>=age_end[17] & age_col<=age_start[17]],na.rm=TRUE),
           mean(var[age_col>=age_end[18] & age_col<=age_start[18]],na.rm=TRUE))
  return(m)
}

#Bambach carbshelf means
var7<-function(var,age_col,age_end,age_start){
  m<-rbind(mean(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
           mean(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
           mean(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
           mean(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
           mean(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
           mean(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
           mean(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
           mean(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
           mean(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
           mean(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
           mean(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
           mean(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
           mean(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
           mean(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
           mean(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
           mean(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE),
           mean(var[age_col>=age_end[17] & age_col<=age_start[17]],na.rm=TRUE),
           mean(var[age_col>=age_end[18] & age_col<=age_start[18]],na.rm=TRUE))
  return(m)
```

```
  }

  ######################
  # Bambach Data Frames #
  ######################

  #Bambach 2myr df
  b2mo<-var1(dat1$dox,dat1$orgage,dat$b2me,dat$b2ms)
  b2sdo<-var2(dat1$dox,dat1$orgage,dat$b2me,dat$b2ms)
  b2sdc<-var3(dat1$dcarb,dat1$orgage,dat$b2me,dat$b2ms)
  b2slsr<-var4(dat1$stron,dat1$orgage,dat$b2me,dat$b2ms)
  b2sdsl<-var5(dat1$sl,dat1$orgage,dat$b2me,dat$b2ms)
  b2mc<-var6(dat1$patc,dat1$orgage,dat$b2me,dat$b2ms)
  b2cm<-var7(dat1$dcarb,dat1$orgage,dat$b2me,dat$b2ms)

  dat.bam2<-
data.frame(cbind(dat$ext_rate_gperg[1:18],b2mo,b2sdo,b2sdc,b2slsr,b2sdsl,b2mc,b2cm))
  row.names(dat.bam2)<-dat$ext_int[1:18]
  colnames(dat.bam2)<-
c("extrate","b2mo","b2sdo","b2sdc","b2slsr","b2sdsl","b2mc","b2cm")

  #Bambach 5myr df
  b5mo<-var1(dat1$dox,dat1$orgage,dat$b2me,dat$b5ms)
  b5sdo<-var2(dat1$dox,dat1$orgage,dat$b2me,dat$b5ms)
  b5sdc<-var3(dat1$dcarb,dat1$orgage,dat$b2me,dat$b5ms)
  b5slsr<-var4(dat1$stron,dat1$orgage,dat$b2me,dat$b5ms)
  b5sdsl<-var5(dat1$sl,dat1$orgage,dat$b2me,dat$b5ms)
  b5mc<-var6(dat1$patc,dat1$orgage,dat$b2me,dat$b5ms)
  b5cm<-var7(dat1$dcarb,dat1$orgage,dat$b2me,dat$b5ms)

  dat.bam5<-
data.frame(cbind(dat$ext_rate_gperg[1:18],b5mo,b5sdo,b5sdc,b5slsr,b5sdsl,b5mc,b5cm))
  row.names(dat.bam5)<-dat$ext_int[1:18]
  colnames(dat.bam5)<-
c("extrate","b5mo","b5sdo","b5sdc","b5slsr","b5sdsl","b5mc","b5cm")

  #Bambach 10myr df
  b10mo<-var1(dat1$dox,dat1$orgage,dat$b2me,dat$b10ms)
  b10sdo<-var2(dat1$dox,dat1$orgage,dat$b2me,dat$b10ms)
  b10sdc<-var3(dat1$dcarb,dat1$orgage,dat$b2me,dat$b10ms)
  b10slsr<-var4(dat1$stron,dat1$orgage,dat$b2me,dat$b10ms)
  b10sdsl<-var5(dat1$sl,dat1$orgage,dat$b2me,dat$b10ms)
  b10mc<-var6(dat1$patc,dat1$orgage,dat$b2me,dat$b10ms)
  b10cm<-var6(dat1$dcarb,dat1$orgage,dat$b2me,dat$b10ms)
```

```
  dat.bam10<-
data.frame(cbind(dat$ext_rate_gperg[1:18],b10mo,b10sdo,b10sdc,b10slsr,b10sdsl,b10mc,b1
0cm))
  row.names(dat.bam10)<-dat$ext_int[1:18]
  colnames(dat.bam10)<-
c("extrate","b10mo","b10sdo","b10sdc","b10slsr","b10sdsl","b10mc","b10cm")

  ####################################

  #functions
  #Alroy ox iso means
  var1<-function(var,age_col,age_end,age_start){
    m<-rbind(mean(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
             mean(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
             mean(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
             mean(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
             mean(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
             mean(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
             mean(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
             mean(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
             mean(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
             mean(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
             mean(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
             mean(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
             mean(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
             mean(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
             mean(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
             mean(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE))
    return(m)
  }

  #Alroy ox iso sd
  var2<-function(var,age_col,age_end,age_start){
    s<-rbind(sd(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
             sd(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
             sd(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
             sd(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
             sd(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
             sd(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
             sd(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
             sd(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
             sd(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
             sd(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
             sd(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
             sd(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
             sd(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
```

```
        sd(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
        sd(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
        sd(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE))
  return(s)
}

#Alroy carb iso sd
var3<-function(var,age_col,age_end,age_start){
  s<-rbind(sd(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
        sd(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
        sd(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
        sd(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
        sd(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
        sd(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
        sd(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
        sd(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
        sd(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
        sd(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
        sd(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
        sd(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
        sd(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
        sd(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
        sd(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
        sd(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE))
  return(s)
}

#Alroy stron iso slope
var4<-function(var,age_col,age_end,age_start){
  {lm1sr<-lm(var~age_col,subset=(age_col>=age_end[1] & age_col<=age_start[1]))
  lm2sr<-lm(var~age_col,subset=(age_col>=age_end[2] & age_col<=age_start[2]))
  lm3sr<-lm(var~age_col,subset=(age_col>=age_end[3] & age_col<=age_start[3]))
  lm4sr<-lm(var~age_col,subset=(age_col>=age_end[4] & age_col<=age_start[4]))
  lm5sr<-lm(var~age_col,subset=(age_col>=age_end[5] & age_col<=age_start[5]))
  lm6sr<-lm(var~age_col,subset=(age_col>=age_end[6] & age_col<=age_start[6]))
  lm7sr<-lm(var~age_col,subset=(age_col>=age_end[7] & age_col<=age_start[7]))
  lm8sr<-lm(var~age_col,subset=(age_col>=age_end[8] & age_col<=age_start[8]))
  lm9sr<-lm(var~age_col,subset=(age_col>=age_end[9] & age_col<=age_start[9]))
  lm10sr<-lm(var~age_col,subset=(age_col>=age_end[10] & age_col<=age_start[10]))
  lm11sr<-lm(var~age_col,subset=(age_col>=age_end[11] & age_col<=age_start[11]))
  lm12sr<-lm(var~age_col,subset=(age_col>=age_end[12] & age_col<=age_start[12]))
  lm13sr<-lm(var~age_col,subset=(age_col>=age_end[13] & age_col<=age_start[13]))
  lm14sr<-lm(var~age_col,subset=(age_col>=age_end[14] & age_col<=age_start[14]))
  lm15sr<-lm(var~age_col,subset=(age_col>=age_end[15] & age_col<=age_start[15]))
  lm16sr<-lm(var~age_col,subset=(age_col>=age_end[16] & age_col<=age_start[16]))}
```

```
   { return(rbind((-1*lm1sr$coefficients[2]),(-1*lm2sr$coefficients[2]),(-
1*lm3sr$coefficients[2]),(-1*lm4sr$coefficients[2]),
                (-1*lm5sr$coefficients[2]),(-1*lm6sr$coefficients[2]),(-
1*lm7sr$coefficients[2]),(-1*lm8sr$coefficients[2]),
                (-1*lm9sr$coefficients[2]),(-1*lm10sr$coefficients[2]),(-
1*lm11sr$coefficients[2]),(-1*lm12sr$coefficients[2]),
                (-1*lm13sr$coefficients[2]),(-1*lm14sr$coefficients[2]),(-
1*lm15sr$coefficients[2]),(-1*lm16sr$coefficients[2])
   ))}
  }

 #Alroy sl sd
 var5<-function(var,age_col,age_end,age_start){
   s<-rbind(sd(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
          sd(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
          sd(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
          sd(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
          sd(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
          sd(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
          sd(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
          sd(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
          sd(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
          sd(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
          sd(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
          sd(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
          sd(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
          sd(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
          sd(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
          sd(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE))
   return(s)
  }

 #Alroy carbshelf means
 var6<-function(var,age_col,age_end,age_start){
   m<-rbind(mean(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
          mean(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
          mean(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
          mean(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
          mean(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
          mean(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
          mean(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
          mean(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
          mean(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
          mean(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
          mean(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
          mean(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
```

```
        mean(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
        mean(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
        mean(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
        mean(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE))
  return(m)
}

#Alroy carbon iso means
var7<-function(var,age_col,age_end,age_start){
  m<-rbind(mean(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
        mean(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
        mean(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
        mean(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
        mean(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
        mean(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
        mean(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
        mean(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
        mean(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
        mean(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
        mean(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
        mean(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
        mean(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
        mean(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
        mean(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
        mean(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE))
  return(m)
}

#####################
# Alroy Data Frames #
#####################

#Alroy 2myr df
a2mo<-var1(dat1$dox,dat1$orgage,dat$a2me,dat$a2ms)
a2sdo<-var2(dat1$dox,dat1$orgage,dat$a2me,dat$a2ms)
a2sdc<-var3(dat1$dcarb,dat1$orgage,dat$a2me,dat$a2ms)
a2slsr<-var4(dat1$stron,dat1$orgage,dat$a2me,dat$a2ms)
a2sdsl<-var5(dat1$sl,dat1$orgage,dat$a2me,dat$a2ms)
a2mc<-var6(dat1$patc,dat1$orgage,dat$a2me,dat$a2ms)
a2cm<-var7(dcarb.sd3.no$dcarb,dcarb.sd3.no$orgage,dat$a2me,dat$a2ms)


dat.alr2<-
data.frame(cbind(dat$ext_rate_gperg[19:34],a2mo,a2sdo,a2sdc,a2slsr,a2sdsl,a2mc,a2cm))
row.names(dat.alr2)<-dat$ext_int[19:34]
colnames(dat.alr2)<-c("extrate","a2mo","a2sdo","a2sdc","a2slsr","a2sdsl","a2mc","a2cm")
```

```
  #Alroy 5myr df
  a5mo<-var1(dat1$dox,dat1$orgage,dat$a2me,dat$a5ms)
  a5sdo<-var2(dat1$dox,dat1$orgage,dat$a2me,dat$a5ms)
  a5sdc<-var3(dat1$dcarb,dat1$orgage,dat$a2me,dat$a5ms)
  a5slsr<-var4(dat1$stron,dat1$orgage,dat$a2me,dat$a5ms)
  a5sdsl<-var5(dat1$sl,dat1$orgage,dat$a2me,dat$a5ms)
  a5mc<-var6(dat1$patc,dat1$orgage,dat$a2me,dat$a5ms)
  a5cm<-var7(dcarb.sd3.no$dcarb,dcarb.sd3.no$orgage,dat$a2me,dat$a5ms)

  dat.alr5<-
data.frame(cbind(dat$ext_rate_gperg[19:34],a5mo,a5sdo,a5sdc,a5slsr,a5sdsl,a5mc,a5cm))
  row.names(dat.alr5)<-dat$ext_int[19:34]
  colnames(dat.alr5)<-c("extrate","a5mo","a5sdo","a5sdc","a5slsr","a5sdsl","a5mc","a5cm")

  #Alroy 10myr df
  a10mo<-var1(dat1$dox,dat1$orgage,dat$a2me,dat$a10ms)
  a10sdo<-var2(dat1$dox,dat1$orgage,dat$a2me,dat$a10ms)
  a10sdc<-var3(dat1$dcarb,dat1$orgage,dat$a2me,dat$a10ms)
  a10slsr<-var4(dat1$stron,dat1$orgage,dat$a2me,dat$a10ms)
  a10sdsl<-var5(dat1$sl,dat1$orgage,dat$a2me,dat$a10ms)
  a10mc<-var6(dat1$patc,dat1$orgage,dat$a2me,dat$a10ms)
  a10cm<-var7(dcarb.sd3.no$dcarb,dcarb.sd3.no$orgage,dat$a2me,dat$a10ms)

  dat.alr10<-
data.frame(cbind(dat$ext_rate_gperg[19:34],a10mo,a10sdo,a10sdc,a10slsr,a10sdsl,a10mc,a1
0cm))
  row.names(dat.alr10)<-dat$ext_int[19:34]
  colnames(dat.alr10)<-
c("extrate","a10mo","a10sdo","a10sdc","a10slsr","a10sdsl","a10mc","a10cm")

  #carbon isotope midpoints
  a2c<-dat$a2ms[1:16]-1
  a5c<-dat$a5ms[1:16]-2.5
  a10c<-dat$a10ms[1:16]-5

  mean.carb<-data.frame(cbind(dat$ext_rate_gperg[19:34],a2c,a5c,a10c,a2cm,a5cm,a10cm))
  row.names(mean.carb)<-dat$ext_int[19:34]
  colnames(mean.carb)<-c("extrate","a2c","a5c","a10c","a2cm","a5cm","a10cm")

  sd.carb<-data.frame(cbind(dat$ext_rate_gperg[19:34],a2c,a5c,a10c,a2sdc,a5sdc,a10sdc))
  row.names(sd.carb)<-dat$ext_int[19:34]
  colnames(sd.carb)<-c("extrate","a2c","a5c","a10c","a2sdc","a5sdc","a10sdc")

  ##################################
  #Within Bin Assessment Dataframe#
```

```
#################################

var8<-function(var,age_col,age_end,age_start){
  s<-rbind(sd(var[age_col>=age_end[1] & age_col<=age_start[1]],na.rm=TRUE),
           sd(var[age_col>=age_end[2] & age_col<=age_start[2]],na.rm=TRUE),
           sd(var[age_col>=age_end[3] & age_col<=age_start[3]],na.rm=TRUE),
           sd(var[age_col>=age_end[4] & age_col<=age_start[4]],na.rm=TRUE),
           sd(var[age_col>=age_end[5] & age_col<=age_start[5]],na.rm=TRUE),
           sd(var[age_col>=age_end[6] & age_col<=age_start[6]],na.rm=TRUE),
           sd(var[age_col>=age_end[7] & age_col<=age_start[7]],na.rm=TRUE),
           sd(var[age_col>=age_end[8] & age_col<=age_start[8]],na.rm=TRUE),
           sd(var[age_col>=age_end[9] & age_col<=age_start[9]],na.rm=TRUE),
           sd(var[age_col>=age_end[10] & age_col<=age_start[10]],na.rm=TRUE),
           sd(var[age_col>=age_end[11] & age_col<=age_start[11]],na.rm=TRUE),
           sd(var[age_col>=age_end[12] & age_col<=age_start[12]],na.rm=TRUE),
           sd(var[age_col>=age_end[13] & age_col<=age_start[13]],na.rm=TRUE),
           sd(var[age_col>=age_end[14] & age_col<=age_start[14]],na.rm=TRUE),
           sd(var[age_col>=age_end[15] & age_col<=age_start[15]],na.rm=TRUE),
           sd(var[age_col>=age_end[16] & age_col<=age_start[16]],na.rm=TRUE))
  return(s)
}

#2myr binned sdc
wib2.sdc<-var8(dat1$dcarb,dat1$orgage,dat$a2me,dat$a2me+2)
wib4.sdc<-var8(dat1$dcarb,dat1$orgage,dat$a2me+2,dat$a2me+4)
wib6.sdc<-var8(dat1$dcarb,dat1$orgage,dat$a2me+4,dat$a2me+6)
wib8.sdc<-var8(dat1$dcarb,dat1$orgage,dat$a2me+6,dat$a2me+8)
wib10.sdc<-var8(dat1$dcarb,dat1$orgage,dat$a2me+8,dat$a2me+10)

wib.sd.carb<-data.frame(cbind(dat$ext_rate_gperg[19:34],wib2.sdc,wib4.sdc,
                        wib6.sdc,wib8.sdc,wib10.sdc))
row.names(wib.sd.carb)<-dat$ext_int[19:34]
colnames(wib.sd.carb)<-c("extrate","0-2","2-4","4-6","6-8","8-10")

#2myr binned sdsl
wib2.sdsl<-var8(dat1$sl,dat1$orgage,dat$a2me,dat$a2me+2)
wib4.sdsl<-var8(dat1$sl,dat1$orgage,dat$a2me+2,dat$a2me+4)
wib6.sdsl<-var8(dat1$sl,dat1$orgage,dat$a2me+4,dat$a2me+6)
wib8.sdsl<-var8(dat1$sl,dat1$orgage,dat$a2me+6,dat$a2me+8)
wib10.sdsl<-var8(dat1$sl,dat1$orgage,dat$a2me+8,dat$a2me+10)

wib.sd.sl<-data.frame(cbind(dat$ext_rate_gperg[19:34],wib2.sdsl,wib4.sdsl,
                        wib6.sdsl,wib8.sdsl,wib10.sdsl))
row.names(wib.sd.sl)<-dat$ext_int[19:34]
colnames(wib.sd.sl)<-c("extrate","0-2","2-4","4-6","6-8","8-10")
```

```
#########################
# Abiotic Summary Stats #
#########################

 count1<-function(var,age_col,age_end,age_start){
  m<-rbind(length(which(!is.na(var[age_col>=age_end[1] & age_col<=age_start[1]]))),
         length(which(!is.na(var[age_col>=age_end[2] & age_col<=age_start[2]]))),
         length(which(!is.na(var[age_col>=age_end[3] & age_col<=age_start[3]]))),
         length(which(!is.na(var[age_col>=age_end[4] & age_col<=age_start[4]]))),
         length(which(!is.na(var[age_col>=age_end[5] & age_col<=age_start[5]]))),
         length(which(!is.na(var[age_col>=age_end[6] & age_col<=age_start[6]]))),
         length(which(!is.na(var[age_col>=age_end[7] & age_col<=age_start[7]]))),
         length(which(!is.na(var[age_col>=age_end[8] & age_col<=age_start[8]]))),
         length(which(!is.na(var[age_col>=age_end[9] & age_col<=age_start[9]]))),
         length(which(!is.na(var[age_col>=age_end[10] & age_col<=age_start[10]]))),
         length(which(!is.na(var[age_col>=age_end[11] & age_col<=age_start[11]]))),
         length(which(!is.na(var[age_col>=age_end[12] & age_col<=age_start[12]]))),
         length(which(!is.na(var[age_col>=age_end[13] & age_col<=age_start[13]]))),
         length(which(!is.na(var[age_col>=age_end[14] & age_col<=age_start[14]]))),
         length(which(!is.na(var[age_col>=age_end[15] & age_col<=age_start[15]]))),
         length(which(!is.na(var[age_col>=age_end[16] & age_col<=age_start[16]]))),
         length(which(!is.na(var[age_col>=age_end[17] & age_col<=age_start[17]]))),
         length(which(!is.na(var[age_col>=age_end[18] & age_col<=age_start[18]]))))
  return(m)
 }

 #Bombach
 #ox iso
 b2ox<-count1(dat1$dox,dat1$orgage,dat$b2me[1:18],dat$b2ms[1:18])
 b2oxd<-describe(b2ox)
 b5ox<-count1(dat1$dox,dat1$orgage,dat$b2me[1:18],dat$b5ms[1:18])
 b5oxd<-describe(b5ox)
 b10ox<-count1(dat1$dox,dat1$orgage,dat$b2me[1:18],dat$b10ms[1:18])
 b10oxd<-describe(b10ox)
 #carb iso
 b2carb<-count1(dat1$dcarb,dat1$orgage,dat$b2me[1:18],dat$b2ms[1:18])
 b2carbd<-describe(b2carb)
 b5carb<-count1(dat1$dcarb,dat1$orgage,dat$b2me[1:18],dat$b5ms[1:18])
 b5carbd<-describe(b5carb)
 b10carb<-count1(dat1$dcarb,dat1$orgage,dat$b2me[1:18],dat$b10ms[1:18])
 b10carbd<-describe(b10carb)
 #stron iso
 b2stron<-count1(dat1$stron,dat1$orgage,dat$b2me[1:18],dat$b2ms[1:18])
 b2strond<-describe(b2stron)
 b5stron<-count1(dat1$stron,dat1$orgage,dat$b2me[1:18],dat$b5ms[1:18])
 b5strond<-describe(b5stron)
```

```
  b10stron<-count1(dat1$stron,dat1$orgage,dat$b2me[1:18],dat$b10ms[1:18])
  b10strond<-describe(b10stron)
  #sl
  b2sl<-count1(dat1$sl,dat1$orgage,dat$b2me[1:18],dat$b2ms[1:18])
  b2sld<-describe(b2sl)
  b5sl<-count1(dat1$sl,dat1$orgage,dat$b2me[1:18],dat$b5ms[1:18])
  b5sld<-describe(b5sl)
  b10sl<-count1(dat1$sl,dat1$orgage,dat$b2me[1:18],dat$b10ms[1:18])
  b10sld<-describe(b10sl)
  #patc
  b2patc<-count1(dat1$patc,dat1$orgage,dat$b2me[1:18],dat$b2ms[1:18])
  b2patcd<-describe(b2patc)
  b5patc<-count1(dat1$patc,dat1$orgage,dat$b2me[1:18],dat$b5ms[1:18])
  b5patcd<-describe(b5patc)
  b10patc<-count1(dat1$patc,dat1$orgage,dat$b2me[1:18],dat$b10ms[1:18])
  b10patcd<-describe(b10patc)


  sumstatb<-
as.data.frame(rbind(b2oxd,b5oxd,b10oxd,b2carbd,b5carbd,b10carbd,b2strond,b5strond,b10st
rond,
                b2sld,b5sld,b10sld,b2patcd,b5patcd,b10patcd))
  row.names(sumstatb)<-c("b2ox","b5ox","b10ox","b2carb","b5carb","b10carb","b2stron",
                   "b5stron","b10stron","b2sl","b5sl","b10sl","b2patc","b5patc","b10patc")

  cpbb<-
as.data.frame(cbind(b2ox,b5ox,b10ox,b2carb,b5carb,b10carb,b2stron,b5stron,b10stron,b2sl,
b5sl,
              b10sl,b2patc,b5patc,b10patc))
  colnames(cpbb)<-
c("b2ox","b5ox","b10ox","b2carb","b5carb","b10carb","b2stron","b5stron","b10stron",
                "b2sl","b5sl","b10sl","b2patc","b5patc","b10patc")
  row.names(cpbb)<-dat$ext_int[1:18]

  ########################################

  count2<-function(var,age_col,age_end,age_start){
   m<-rbind(length(which(!is.na(var[age_col>=age_end[1] & age_col<=age_start[1]]))),
          length(which(!is.na(var[age_col>=age_end[2] & age_col<=age_start[2]]))),
          length(which(!is.na(var[age_col>=age_end[3] & age_col<=age_start[3]]))),
          length(which(!is.na(var[age_col>=age_end[4] & age_col<=age_start[4]]))),
          length(which(!is.na(var[age_col>=age_end[5] & age_col<=age_start[5]]))),
          length(which(!is.na(var[age_col>=age_end[6] & age_col<=age_start[6]]))),
          length(which(!is.na(var[age_col>=age_end[7] & age_col<=age_start[7]]))),
          length(which(!is.na(var[age_col>=age_end[8] & age_col<=age_start[8]]))),
          length(which(!is.na(var[age_col>=age_end[9] & age_col<=age_start[9]]))),
```

```
        length(which(!is.na(var[age_col>=age_end[10] & age_col<=age_start[10]]))),
        length(which(!is.na(var[age_col>=age_end[11] & age_col<=age_start[11]]))),
        length(which(!is.na(var[age_col>=age_end[12] & age_col<=age_start[12]]))),
        length(which(!is.na(var[age_col>=age_end[13] & age_col<=age_start[13]]))),
        length(which(!is.na(var[age_col>=age_end[14] & age_col<=age_start[14]]))),
        length(which(!is.na(var[age_col>=age_end[15] & age_col<=age_start[15]]))),
        length(which(!is.na(var[age_col>=age_end[16] & age_col<=age_start[16]]))))
  return(m)
}

#Alroy
#ox iso
a2ox<-count2(dat1$dox,dat1$orgage,dat$a2me[1:16],dat$a2ms[1:16])
a2oxd<-describe(a2ox)
a5ox<-count2(dat1$dox,dat1$orgage,dat$a2me[1:16],dat$a5ms[1:16])
a5oxd<-describe(a5ox)
a10ox<-count2(dat1$dox,dat1$orgage,dat$a2me[1:16],dat$a10ms[1:16])
a10oxd<-describe(a10ox)
#carb iso
a2carb<-count2(dat1$dcarb,dat1$orgage,dat$a2me[1:16],dat$a2ms[1:16])
a2carbd<-describe(a2carb)
a5carb<-count2(dat1$dcarb,dat1$orgage,dat$a2me[1:16],dat$a5ms[1:16])
a5carbd<-describe(a5carb)
a10carb<-count2(dat1$dcarb,dat1$orgage,dat$a2me[1:16],dat$a10ms[1:16])
a10carbd<-describe(a10carb)
#stron iso
a2stron<-count2(dat1$stron,dat1$orgage,dat$a2me[1:16],dat$a2ms[1:16])
a2strond<-describe(a2stron)
a5stron<-count2(dat1$stron,dat1$orgage,dat$a2me[1:16],dat$a5ms[1:16])
a5strond<-describe(a5stron)
a10stron<-count2(dat1$stron,dat1$orgage,dat$a2me[1:16],dat$a10ms[1:16])
a10strond<-describe(a10stron)
#sl
a2sl<-count2(dat1$sl,dat1$orgage,dat$a2me[1:16],dat$a2ms[1:16])
a2sld<-describe(a2sl)
a5sl<-count2(dat1$sl,dat1$orgage,dat$a2me[1:16],dat$a5ms[1:16])
a5sld<-describe(a5sl)
a10sl<-count2(dat1$sl,dat1$orgage,dat$a2me[1:16],dat$a10ms[1:16])
a10sld<-describe(a10sl)
#patc
a2patc<-count2(dat1$patc,dat1$orgage,dat$a2me[1:16],dat$a2ms[1:16])
a2patcd<-describe(a2patc)
a5patc<-count2(dat1$patc,dat1$orgage,dat$a2me[1:16],dat$a5ms[1:16])
a5patcd<-describe(a5patc)
a10patc<-count2(dat1$patc,dat1$orgage,dat$a2me[1:16],dat$a10ms[1:16])
a10patcd<-describe(a10patc)
```

```
  sumstata<-
as.data.frame(rbind(a2oxd,a5oxd,a10oxd,a2carbd,a5carbd,a10carbd,a2strond,a5strond,a10str
ond,
                 a2sld,a5sld,a10sld,a2patcd,a5patcd,a10patcd))
  row.names(sumstata)<-c("a2ox","a5ox","a10ox","a2carb","a5carb","a10carb","a2stron",
                        "a5stron","a10stron","a2sl","a5sl","a10sl","a2patc","a5patc","a10patc")

  cpaa<-
as.data.frame(cbind(a2ox,a5ox,a10ox,a2carb,a5carb,a10carb,a2stron,a5stron,a10stron,a2sl,a5
sl,
             a10sl,a2patc,a5patc,a10patc))
  colnames(cpaa)<-
c("a2ox","a5ox","a10ox","a2carb","a5carb","a10carb","a2stron","a5stron","a10stron",
                   "a2sl","a5sl","a10sl","a2patc","a5patc","a10patc")
  row.names(cpaa)<-dat$ext_int[19:34]

  sumstatb
  cpbb
  sumstata
  cpaa

#Extinction rate vs time plots
plot(dat.alr2$extrate, xlab="",xaxt = "n",main="Extincton Rate vs Interval
(Time)",ylab="Extinction Rate")
axis(side=1,labels=F)
axis(1, at=1:16, labels=rownames(dat.alr2), las=2)

#Environmental variable boxplots
boxplot(dat1[,-c(1,5,6)])
boxplot(dat1[,5])

#dat.bam2

    p <- ggpairs(dat.bam2[,c(1:6)],title="Scatter Plot, Correlation, and Distribution Matrix")
    print(p)

    #Bambach model 1
    bmod1<-lm(extrate~.-b2mc-b2cm,data=dat.bam2,na.action=na.omit)
    summary(bmod1)
    p1<-glance(bmod1)$p.value
    p.adjust(p1,"bonferroni",5)
    #Predictive R2
    ##model validation using PRESS statistic
    #removes each observation, refits model,
```

```
#calculates the out-of-sample predicted
#value for each omitted observation, Press
#is calculated as the sum of squares of all
#the resulting prediction errors.
#lowest values of PRESS indicate the best structures

bmod1at<-drop1(bmod1, .~., test="F")
bmod1ts<-sum(bmod1at[2:6,2],bmod1at[1,3])
bmod1pr2<-(1-(PRESS(bmod1)/bmod1ts))*100
bmod1pr2

#diagnostic plots
        plot(bmod1,which=c(1,4))
  #residuals v.s fitted values
  plot(dat.bam2$b2mo[-c(11,15)], bmod1$residuals,main="Res. vs Mean 18O")
  abline(h = 0, col = "gray75")
  plot(dat.bam2$b2sdo[-c(11,15)], bmod1$residuals,main="Res. vs Stdev 18O")
  abline(h = 0, col = "gray75")
  plot(dat.bam2$b2sdc[-c(11,15)], bmod1$residuals,main="Res. vs Stdev 13C")
  abline(h = 0, col = "gray75")
  plot(dat.bam2$b2slsr[-c(11,15)], bmod1$residuals,main="Res. vs Slope 87Sr/86Sr")
  abline(h = 0, col = "gray75")
  plot(dat.bam2$b2sdsl[-c(11,15)], bmod1$residuals,main="Res. vs Stdev Sea Level")
  abline(h = 0, col = "gray75")
  qqPlot(bmod1$residuals, las = 1, main="QQ Plot")
  std <- studres(bmod1) #standardized residuals
  plot(std,main="Standardized Residuals vs Observation Number")
  lev<-hatvalues(bmod1)
  plot(lev,main="LeverAge vs Observation Number")


#table of diagnostic info
as.data<-as.matrix(cbind(bmod1$fitted.values,bmod1$residuals,std,lev))
colnames(as.data)<-c("fit","res","std","lev")
as.data
p=6
n=15
2*p/n
3*p/n

#normality

  std.res<-(rstandard(bmod1))
  h<-hist(std.res)
  xfit<-seq(min(std.res),max(std.res),length=40)
  yfit<-dnorm(xfit,mean=mean(std.res),sd=sd(std.res))
```

```
    yfit <- yfit*diff(h$mids[1:2])*length(std.res)
    lines(xfit, yfit, col="blue", lwd=2)

  describe(std.res)
  shapiro.test(rstandard(bmod1))

  #error variance
  bptest(extrate~.,data=dat.bam2,studentize=FALSE)
  #autocor
  checkresiduals(bmod1,test="BG",lag=1)
  bgtest(bmod1,type="F")

  #multicollinearity
  vif(bmod1)

  #outliers
  outlierTest(bmod1)
#dat.bam5

  p <- ggpairs(dat.bam5[,c(1:6)],title="Scatter Plot, Correlation, and Distribution Matrix")
  print(p)

  #Bambach model 2
  bmod2<-lm(extrate~.-b5mc-b5cm,data=dat.bam5,na.action=na.omit)
  summary(bmod2)
  p2<-glance(bmod2)$p.value
  p.adjust(p2,"bonferroni",5)
  #Predictive R2
  bmod2at<-drop1(bmod2, .~., test="F")
  bmod2ts<-sum(bmod2at[2:6,2],bmod2at[1,3])
  bmod2pr2<-(1-(PRESS(bmod2)/bmod2ts))*100
  bmod2pr2

  #diagnostic plots
    plot(bmod2,which=c(1,4))
    #residuals v.s fitted values
    plot(dat.bam5$b5mo, bmod2$residuals,main="Res. vs Mean 18O")
    abline(h = 0, col = "gray75")
    plot(dat.bam5$b5sdo, bmod2$residuals,main="Res. vs Stdev 18O")
    abline(h = 0, col = "gray75")
    plot(dat.bam5$b5sdc, bmod2$residuals,main="Res. vs Stdev 13C")
    abline(h = 0, col = "gray75")
    plot(dat.bam5$b5slsr, bmod2$residuals,main="Res. vs Slope 87Sr/86Sr")
    abline(h = 0, col = "gray75")
    plot(dat.bam5$b5sdsl, bmod2$residuals,main="Res. vs Stdev Sea Level")
    abline(h = 0, col = "gray75")
```

```
    qqPlot(bmod2$residuals, las = 1, main="QQ Plot")
    std <- studres(bmod2) #standardized residuals
    plot(std,main="Standardized Residuals vs Observation Number")
    lev<-hatvalues(bmod2)
    plot(lev,main="LeverAge vs Observation Number")

  #table of diagnostic info
  as.data<-as.matrix(cbind(bmod2$fitted.values,bmod2$residuals,std,lev))
  colnames(as.data)<-c("fit","res","std","lev")
  as.data
  p=6
  n=18
  2*p/n
  3*p/n

  #normality

    std.res<-(rstandard(bmod2))
    h<-hist(std.res)
    xfit<-seq(min(std.res),max(std.res),length=40)
    yfit<-dnorm(xfit,mean=mean(std.res),sd=sd(std.res))
    yfit <- yfit*diff(h$mids[1:2])*length(std.res)
    lines(xfit, yfit, col="blue", lwd=2)

  describe(std.res)
  shapiro.test(rstandard(bmod2))

  #error variance
  bptest(extrate~.,data=dat.bam5,studentize=FALSE)
  #autocor
  checkresiduals(bmod2,test="BG",lag=1)
  bgtest(bmod2,type="F")

  #multicollinearity
  vif(bmod2)

  #outliers
  outlierTest(bmod2)

#dat.bam10

  p <- ggpairs(dat.bam10[,c(1:6)],title="Scatter Plot, Correlation, and Distribution Matrix")
  print(p)

  #Bambach model 3
  bmod3<-lm(extrate~.-b10mc-b10cm,data=dat.bam10,na.action=na.omit)
```

```
summary(bmod3)
p3<-glance(bmod3)$p.value
p.adjust(p3,"bonferroni",5)
#Predictive R2
bmod3at<-drop1(bmod3, .~., test="F")
bmod3ts<-sum(bmod3at[2:6,2],bmod3at[1,3])
bmod3pr2<-(1-(PRESS(bmod3)/bmod3ts))*100
bmod3pr2

#diagnostic plots
  plot(bmod3,which=c(1,4))
  #residuals v.s fitted values
  plot(dat.bam10$b10mo, bmod3$residuals,main="Res. vs Mean 18O")
  abline(h = 0, col = "gray75")
  plot(dat.bam10$b10sdo, bmod3$residuals,main="Res. vs Stdev 18O")
  abline(h = 0, col = "gray75")
  plot(dat.bam10$b10sdc, bmod3$residuals,main="Res. vs Stdev 13C")
  abline(h = 0, col = "gray75")
  plot(dat.bam10$b10slsr, bmod3$residuals,main="Res. vs Slope 87Sr/86Sr")
  abline(h = 0, col = "gray75")
  plot(dat.bam10$b10sdsl, bmod3$residuals,main="Res. vs Stdev Sea Level")
  abline(h = 0, col = "gray75")
  qqPlot(bmod3$residuals, las = 1, main="QQ Plot")
  std <- studres(bmod3) #standardized residuals
  plot(std,main="Standardized Residuals vs Observation Number")
  lev<-hatvalues(bmod3)
  plot(lev,main="LeverAge vs Observation Number")

#table of diagnostic info
as.data<-as.matrix(cbind(bmod3$fitted.values,bmod3$residuals,std,lev))
colnames(as.data)<-c("fit","res","std","lev")
as.data
p=6
n=18
2*p/n
3*p/n

#normality
  std.res<-(rstandard(bmod3))
  h<-hist(std.res)
  xfit<-seq(min(std.res),max(std.res),length=40)
  yfit<-dnorm(xfit,mean=mean(std.res),sd=sd(std.res))
  yfit <- yfit*diff(h$mids[1:2])*length(std.res)
  lines(xfit, yfit, col="blue", lwd=2)
describe(std.res)
shapiro.test(rstandard(bmod3))
```

```
  #error variance
  bptest(extrate~.,data=dat.bam10,studentize=FALSE)
  #autocor
  checkresiduals(bmod3,test="BG",lag=1)
  bgtest(bmod3,type="F")

  #multicollinearity
  vif(bmod3)

  #outliers
  outlierTest(bmod3)

#dat.alr2
p <- ggpairs(dat.alr2[,c(1:6)],title="Scatter Plot, Correlation, and Distribution Matrix")
print(p)

#Alroy model 1
mod1<-lm(extrate~.,data=dat.alr2[,c(1:6)],na.action=na.omit)
mod1sum<-summary(mod1)
mod1sum
p1<-glance(mod1)$p.value
p1ad<-p.adjust(p1,"bonferroni",6)
p1ad
#Predictive R2
mod1at<-drop1(mod1, .~., test="F")
mod1ts<-sum(mod1at[2:6,2],mod1at[1,3])
mod1pr2<-(1-(PRESS(mod1)/mod1ts))
mod1pr2

#diagnostic plots
  plot(mod1,which=c(1,4))
  #residuals v.s fitted values
  plot(dat.alr2$a2mo, mod1$residuals,main="Res. vs Mean 18O")
  abline(h = 0, col = "gray75")
  plot(dat.alr2$a2sdo, mod1$residuals,main="Res. vs Stdev 18O")
  abline(h = 0, col = "gray75")
  plot(dat.alr2$a2sdc, mod1$residuals,main="Res. vs Stdev 13C")
  abline(h = 0, col = "gray75")
  plot(dat.alr2$a2slsr, mod1$residuals,main="Res. vs Slope 87Sr/86Sr")
  abline(h = 0, col = "gray75")
  plot(dat.alr2$a2sdsl, mod1$residuals,main="Res. vs Stdev Sea Level")
  abline(h = 0, col = "gray75")
  qqPlot(mod1$residuals, las = 1, main="QQ Plot")
  std <- studres(mod1) #standardized residuals
  plot(std,main="Standardized Residuals vs Observation Number")
```

```
  lev<-hatvalues(mod1)
  plot(lev,main="LeverAge vs Observation Number")

#table of diagnostic info
as.data<-as.matrix(cbind(mod1$fitted.values,mod1$residuals,std,lev))
colnames(as.data)<-c("fit","res","std","lev")
as.data
p=6
n=16
2*p/n
3*p/n

#normality
  std.res<-(rstandard(mod1))
  h<-hist(std.res)
  xfit<-seq(min(std.res),max(std.res),length=40)
  yfit<-dnorm(xfit,mean=mean(std.res),sd=sd(std.res))
  yfit <- yfit*diff(h$mids[1:2])*length(std.res)
  lines(xfit, yfit, col="blue", lwd=2)
describe(std.res)
shapiro.test(rstandard(mod1))

#error variance
bptest(extrate~.,data=dat.alr2,studentize=FALSE)
#autocor
checkresiduals(mod1,test="BG",lag=1)
bgtest(mod1,type="F")

#multicollinearity
vif(mod1)

#Bonferroni outlier test
outlierTest(mod1)

#interactions
#sdc*sdsl
plot(dat.alr2[,4]*dat.alr2[,6],mod1$residuals,main="std13c*sdsl")

#dat.alr5

p <- ggpairs(dat.alr5[,c(1:6)],title="Scatter Plot, Correlation, and Distribution Matrix")
print(p)

#Alroy model 2
mod2<-lm(extrate~.,data=dat.alr5[,c(1:6)],na.action=na.omit)
```

```
  mod2<-lm(extrate~.+a5sdo*a5sdc+a5sdc*a5sdsl+a5sdo*a5sdsl,data=dat.alr5[,c(1:6)],na.action=na.omit)
  mod2<-lm(extrate~.+a5sdo*a5sdsl,data=dat.alr5[,c(1:6)],na.action=na.omit)
  mod2sum<-summary(mod2)
  mod2sum
  p2<-glance(mod2)$p.value
  p2ad<-p.adjust(p2,"bonferroni",6)
  p2ad
  #SSS table
  options(contrasts = c("contr.sum","contr.poly"))
  drop1(mod2, .~., test="F")
  #Predictive R2
  mod2at<-drop1(mod2, .~., test="F")
  mod2ts<-sum(mod2at[2:7,2],mod2at[1,3])
  mod2pr2<-(1-(PRESS(mod2)/mod2ts))
  mod2pr2

  #diagnostic plots

    plot(mod2,which=c(1,4))
    #residuals v.s fitted values
    plot(dat.alr5$a5mo,mod2$residuals,main="Res. vs Mean 18O")
    abline(h = 0, col = "gray75")
    plot(dat.alr5$a5sdo,mod2$residuals,main="Res. vs Stdev 18O")
    abline(h = 0, col = "gray75")
    plot(dat.alr5$a5sdc,mod2$residuals,main="Res. vs Stdev 13C")
    abline(h = 0, col = "gray75")
    plot(dat.alr5$a5slsr,mod2$residuals,main="Res. vs Slope 87Sr/86Sr")
    abline(h = 0, col = "gray75")
    plot(dat.alr5$a5sdsl,mod2$residuals,main="Res. vs Stdev Sea Level")
    abline(h = 0, col = "gray75")
    qqPlot(mod2$residuals, las = 1, main="QQ Plot")
    std <- studres(mod2) #standardized residuals
    plot(std,main="Standardized Residuals vs Observation Number")
    lev<-hatvalues(mod2)
    plot(lev,main="Leverage vs Observation Number")

  #table of diagnostic info
  as.data<-as.matrix(cbind(mod2$fitted.values,mod2$residuals,std,lev))
  colnames(as.data)<-c("fit","res","std","lev")
  as.data
  p=7
  n=16
  2*p/n
  3*p/n
```

```
  #normality
    std.res<-(rstandard(mod2))
    h<-hist(std.res)
    xfit<-seq(min(std.res),max(std.res),length=40)
    yfit<-dnorm(xfit,mean=mean(std.res),sd=sd(std.res))
    yfit <- yfit*diff(h$mids[1:2])*length(std.res)
    lines(xfit, yfit, col="blue", lwd=2)
  describe(std.res)
  shapiro.test(rstandard(mod2))

  #error variance
  bptest(extrate~.,data=dat.alr5,studentize=TRUE)
  #autocor
  checkresiduals(mod2,test="BG",lag=1)
  bgtest(mod2,type="F")

  #multicollinearity
  vif(mod2)

  #outliers
  outlierTest(mod2)

  #interactions
  #sdo*sdc
  plot(dat.alr5[,3]*dat.alr5[,4],mod2$residuals,main="std18o*std13c")
  #sdc*sdsl
  plot(dat.alr5[,4]*dat.alr5[,6],mod2$residuals,main="std13c*sdsl")
  #sdo*sdsl
  plot(dat.alr10[,3]*dat.alr10[,6],mod2$residuals,main="std18o*sdsl")


#################################################################################
###
  ###10 My intervals

  #dat.alr10

  p <- ggpairs(dat.alr10[,c(1:6)],title="Scatter Plot, Correlation, and Distribution Matrix")
  print(p)

  #Alroy model 3
  mod3<-lm(extrate~.,data=dat.alr10[,c(1:6)],na.action=na.omit)
  mod3sum<-summary(mod3)
  mod3sum
  p3<-glance(mod3)$p.value
```

```
p3ad<-p.adjust(p3,"bonferroni",5)
p3ad
#SSS table
options(contrasts = c("contr.sum","contr.poly"))
drop1(mod3, .~., test="F")
#Predictive R2
mod3at<-drop1(mod3, .~., test="F")
mod3ts<-sum(mod3at[2:6,2],mod3at[1,3])
mod3pr2<-(1-(PRESS(mod3)/mod3ts))
mod3pr2

#diagnostic plots

  plot(mod3,which=c(1,4))
  #residuals v.s fitted values
  plot(dat.alr10$a10mo, mod3$residuals,main="Res. vs Mean 18O")
  abline(h = 0, col = "gray75")
  plot(dat.alr10$a10sdo, mod3$residuals,main="Res. vs Stdev 18O")
  abline(h = 0, col = "gray75")
  plot(dat.alr10$a10sdc, mod3$residuals,main="Res. vs Stdev 13C")
  abline(h = 0, col = "gray75")
  plot(dat.alr10$a10slsr, mod3$residuals,main="Res. vs Slope 87Sr/86Sr")
  abline(h = 0, col = "gray75")
  plot(dat.alr10$a10sdsl, mod3$residuals,main="Res. vs Stdev Sea Level")
  abline(h = 0, col = "gray75")
  qqPlot(mod3$residuals, las = 1, main="QQ Plot")
  std <- studres(mod3) #standardized residuals
  plot(std,main="Standardized Residuals vs Observation Number")
  lev<-hatvalues(mod3)
  plot(lev,main="Leverage vs Observation Number")

#table of diagnostic info
as.data<-as.matrix(cbind(mod3$fitted.values,mod3$residuals,std,lev))
colnames(as.data)<-c("fit","res","std","lev")
as.data
p=6
n=16
2*p/n
3*p/n

#normality
  std.res<-(rstandard(mod3))
  h<-hist(std.res)
  xfit<-seq(min(std.res),max(std.res),length=40)
  yfit<-dnorm(xfit,mean=mean(std.res),sd=sd(std.res))
  yfit <- yfit*diff(h$mids[1:2])*length(std.res)
```

```
   lines(xfit, yfit, col="blue", lwd=2)
  describe(std.res)
  shapiro.test(rstandard(mod3))

  #error variance
  bptest(extrate~.,data=dat.alr10,studentize=FALSE)
  #autocor
  checkresiduals(mod3,test="BG",lag=1)
  bgtest(mod3,type="F")

  #multicollinearity
  vif(mod3)

  #outliers
  outlierTest(mod3)

  ####################
#Type 3 ANOVA Table#
####################


  a2myr.vp<-(data.frame(mod1sum[["coefficients"]][,"Pr(>|t|)"][2:6]))
  a2myr.p<-rbind(p1ad,a2myr.vp)
  anova.a2myr<-a2myr.p
  a2r2<-data.frame(mod1sum$r.squared)
  a2ir2<-calc.relimp(mod1,type="lmg")#fit considering all other variables in the model
  a2ir2<-as.data.frame(a2ir2@lmg)
  colnames(a2ir2)<-"a2ir2"
  a2r2<-InsertRow(a2ir2,a2r2,1)
  anova.a2myr$a2r2<-a2r2
  a2ar2<-data.frame(mod1sum$adj.r.squared)
  length(a2ar2)<-nrow(a2myr.p)
  anova.a2myr$a2ar2<-a2ar2
  mod1pr2<-data.frame(mod1pr2)
  length(mod1pr2)<-nrow(a2myr.p)
  anova.a2myr$mod1pr2<-mod1pr2
  colnames(anova.a2myr)<-c("P-Value","Multiple R2 (subsequent values are relative
R2)","Adjusted R2","Predicted R2")
  row.names(anova.a2myr)<-c("2 Myr-prior Model","Average Temperature","Temperature
Stability","Carbon Cycle Stability","Continental Weathering Rates","Habitat Stability")
  colnames(anova.a2myr[2])<-"Multiple R2 (subsequent values are relative R2)"
  anova.a2myr<-as.matrix.data.frame(anova.a2myr)

  a5myr.vp<-(data.frame(mod2sum[["coefficients"]][,"Pr(>|t|)"][2:7]))
  a5myr.p<-rbind(p2ad,a5myr.vp)
  anova.a5myr<-a5myr.p
```

```
  a5r2<-data.frame(mod2sum$r.squared)
  a5ir2<-calc.relimp(mod2,type="lmg")#fit considering all other variables in the model
  a5ir2<-as.data.frame(a5ir2@lmg)
  colnames(a5ir2)<-"a5ir2"
  a5r2<-InsertRow(a5ir2,a5r2,1)
  anova.a5myr$a5r2<-a5r2
  a5ar2<-data.frame(mod2sum$adj.r.squared)
  length(a5ar2)<-nrow(a5myr.p)
  anova.a5myr$a5ar2<-a5ar2
  mod1pr2<-data.frame(mod2pr2)
  length(mod2pr2)<-nrow(a5myr.p)
  anova.a5myr$mod2pr2<-mod2pr2
  colnames(anova.a5myr)<-c("P-Value","Multiple R2 and Relative Importance of
Variables","Adjusted R2","Predicted R2")
  row.names(anova.a5myr)<-c("5 Myr-prior Model","Average Temperature","Temperature
Stability","Carbon Cycle Stability","Continental Weathering Rates","Habitat
Stability","Temperature-Sea Level Interaction")
  anova.a5myr<-as.matrix.data.frame(anova.a5myr)

  a10myr.vp<-(data.frame(mod3sum[["coefficients"]][,"Pr(>|t|)"][2:6]))
  a10myr.p<-rbind(p3ad,a10myr.vp)
  anova.a10myr<-a10myr.p
  a10r2<-data.frame(mod3sum$r.squared)
  a10ir2<-calc.relimp(mod3,type="lmg")#fit considering all other variables in the model
  a10ir2<-as.data.frame(a10ir2@lmg)
  colnames(a10ir2)<-"a10ir2"
  a10r2<-InsertRow(a10ir2,a10r2,1)
  anova.a10myr$a10r2<-a10r2
  a10ar2<-data.frame(mod3sum$adj.r.squared)
  length(a10ar2)<-nrow(a10myr.p)
  anova.a10myr$a10ar2<-a10ar2
  mod3pr2<-data.frame(mod3pr2)
  length(mod3pr2)<-nrow(a10myr.p)
  anova.a10myr$mod3pr2<-mod3pr2
  colnames(anova.a10myr)<-c("P-Value","Multiple R2 (subsequent values are relative
R2)","Adjusted R2","Predicted R2")
  row.names(anova.a10myr)<-c("10 Myr-prior Model","Average Temperature","Temperature
Stability","Carbon Cycle Stability","Continental Weathering Rates","Habitat Stability")
  anova.a10myr<-as.matrix.data.frame(anova.a10myr)

##################
#Model assessment#
##################

#partialled out predictor effect
```

```
d2<-cbind(dat.alr2,(rownames(dat.alr2)))
colnames(d2)[9] <- "extint"
color_pallet_function <- colorRampPalette(
  colors = c("blue", "brown", "purple"),
  space = "Lab" # Option used when colors do not represent a quantitative scale
)
num_colors <- levels(d2$extint)
ext_int_color_colors <- color_pallet_function(num_colors)

  peplot <- function(mod,var,ci=.95, plot_points = "n",
                  xlab=var,ylab=names(mod[12]$model)[1],
                  main="",
                  pe_lty=1,pe_lwd=3,pe_col="black",
                  ci_lty=1,ci_lwd=1,ci_col="black",
                  pch_col="black",pch_ty=19,
                  pe_cex=1,pe_font=1,
                  ylim=c(min(pred[,"lwr"]),max(pred[,"upr"]))){
  modDat <- mod[12]$model
  modDat1 <- modDat[,-1]
  modDat2 <- modDat[,which(names(modDat)!=var)]
  x <- resid(lm(modDat1[,var] ~., data=modDat1[,which(names(modDat1)!=var)]))
  y <- resid(lm(modDat2[,1] ~ ., modDat2[,-1]))
  plot(x,y,type=plot_points,xlab=xlab,ylab=ylab,
       ylim=ylim,col=pch_col,pch=pch_ty,cex=pe_cex,
       main=main,font=1)
  part <- lm(y~x)
  wx <- par("usr")[1:2]
  new.x <- seq(wx[1],wx[2],len=100)
  conf <- predict(part, new=data.frame(x=new.x), interval="conf",
                 level = ci)
  lines(new.x,conf[,"fit"],lwd=pe_lwd,lty=pe_lty,col=pe_col)
  lines(new.x,conf[,"lwr"],lwd=ci_lwd,lty=ci_lty,col="steelblue")
  lines(new.x,conf[,"upr"],lwd=ci_lwd,lty=ci_lty,col="steelblue")
  pred <- predict(part, new=data.frame(x=new.x), interval="prediction",
                 level = ci)
  lines(new.x,pred[,"fit"],lwd=pe_lwd,lty=pe_lty,col=pe_col)
  lines(new.x,pred[,"lwr"],lwd=ci_lwd,lty=ci_lty,col="darkred")
  lines(new.x,pred[,"upr"],lwd=ci_lwd,lty=ci_lty,col="darkred")


}
peplot.l <- function(mod,var,ci=.95, plot_points = "n",
                    xlab=var,ylab=names(mod[12]$model)[1],
                    main="",
                    pe_lty=1,pe_lwd=3,pe_col="black",
                    ci_lty=1,ci_lwd=1,ci_col="black",
```

```
                  pch_col="black",pch_ty=19,
                  pe_cex=1,
                  ylim=c(min(pred[,"lwr"]),max(pred[,"upr"]))){
  modDat <- mod[12]$model
  modDat1 <- modDat[,-1]
  modDat2 <- modDat[,which(names(modDat)!=var)]
  x <- resid(lm(modDat1[,var] ~., data=modDat1[,which(names(modDat1)!=var)]))
  y <- resid(lm(modDat2[,1] ~ ., modDat2[,-1]))
  plot(x,y,type=plot_points,xlab=xlab,ylab=ylab,
      ylim=ylim,col=pch_col,pch=pch_ty,cex=pe_cex,
      main=main)
  legend("bottomright",
        legend = d2$extint,
        col =ext_int_color_colors,
        pch = c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
        bty = "y",pt.cex = 0.79,cex = 0.29,text.col = "black",
        horiz = F ,inset =F, ncol=2,title="Extinction Interval")
  part <- lm(y~x)
  wx <- par("usr")[1:2]
  new.x <- seq(wx[1],wx[2],len=100)
  conf <- predict(part, new=data.frame(x=new.x), interval="conf",
                level = ci)
  lines(new.x,conf[,"fit"],lwd=pe_lwd,lty=pe_lty,col=pe_col)
  lines(new.x,conf[,"lwr"],lwd=ci_lwd,lty=ci_lty,col="steelblue")
  lines(new.x,conf[,"upr"],lwd=ci_lwd,lty=ci_lty,col="steelblue")
  pred <- predict(part, new=data.frame(x=new.x), interval="prediction",
                level = ci)
  lines(new.x,pred[,"fit"],lwd=pe_lwd,lty=pe_lty,col=pe_col)
  lines(new.x,pred[,"lwr"],lwd=ci_lwd,lty=ci_lty,col="darkred")
  lines(new.x,pred[,"upr"],lwd=ci_lwd,lty=ci_lty,col="darkred")


}

#2myr

par(mfrow=c(3,2),family="serif",bty="l")
peplot(mod1,"a2mo",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
      pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
      ci_lty=2,xlab=expression("Mean Temperature (r"[i]*")"),
      ylab=expression("Extinction Rate (r"[i]*")"),
      pe_cex=2,pe_font=2)
title(main="A.",adj=0,family="serif",font=2)
peplot(mod1,"a2sdo",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
      pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
      ci_lty=2,xlab=expression("Temperature Instability (r"[i]*")"),
```

```
        ylab=expression("Extinction Rate (r"[i]*")"),
        pe_cex=2)
title(main="B.",adj=0,family="serif",font=2)
peplot(mod1,"a2sdc",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
        pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
        ci_lty=2,xlab=expression("Carbon Cycle Instability (r"[i]*")"),
        ylab=expression("Extinction Rate (r"[i]*")"),
        pe_cex=2)
title(main="C.",adj=0,family="serif",font=2)
peplot(mod1,"a2slsr",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
        pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
        ci_lty=2,xlab=expression("Continental Weathering Rates (r"[i]*")"),
        ylab=expression("Extinction Rate (r"[i]*")"),
        pe_cex=2)
title(main="D.",adj=0,family="serif",font=2)
peplot(mod1,"a2sdsl",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
        pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
        ci_lty=2,xlab=expression("Habitat Instability (r"[i]*")"),
        ylab=expression("Extinction Rate (r"[i]*")"),
        pe_cex=2)
title(main="E.",adj=0,family="serif",font=2)

#5myr

  par(mfrow=c(3,2),family="serif",bty="l")
  peplot(mod2,"a5mo",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
          pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
          ci_lty=2,xlab=expression("Mean Temperature (r"[i]*")"),
          ylab=expression("Extinction Rate (r"[i]*")"),
          pe_cex=2)
  title(main="A.",adj=0,family="serif",font=2)
  peplot(mod2,"a5sdo",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
          pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
          ci_lty=2,xlab=expression("Temperature Instability (r"[i]*")"),
          ylab=expression("Extinction Rate (r"[i]*")"),
          pe_cex=2)
  title(main="B.",adj=0,family="serif",font=2)
  peplot(mod2,"a5sdc",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
          pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
          ci_lty=2,xlab=expression("Carbon Cycle Instability (r"[i]*")"),
          ylab=expression("Extinction Rate (r"[i]*")"),
          pe_cex=2)
  title(main="C.",adj=0,family="serif",font=2)
  peplot(mod2,"a5slsr",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
          pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
          ci_lty=2,xlab=expression("Continental Weathering Rates (r"[i]*")"),
```

```
      ylab=expression("Extinction Rate (r"[i]*")"),
      pe_cex=2)
title(main="D.",adj=0,family="serif",font=2)
peplot(mod2,"a5sdsl",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
      pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
      ci_lty=2,xlab=expression("Sea Level Instability (r"[i]*")"),
      ylab=expression("Extinction Rate (r"[i]*")"),
      pe_cex=2)
title(main="E.",adj=0,family="serif",font=2)

par(mfrow=c(3,2),family="serif",bty="l")
peplot(mod3,"a10mo",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
      pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
      ci_lty=2,xlab=expression("Mean Temperature (r"[i]*")"),
      ylab=expression("Extinction Rate (r"[i]*")"),
      pe_cex=2)
title(main="A.",adj=0,family="serif",font=2)
peplot(mod3,"a10sdo",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
      pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
      ci_lty=2,xlab=expression("Temperature Instability (r"[i]*")"),
      ylab=expression("Extinction Rate (r"[i]*")"),
      pe_cex=2)
title(main="B.",adj=0,family="serif",font=2)
peplot(mod3,"a10sdc",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
      pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
      ci_lty=2,xlab=expression("Carbon Cycle Instability (r"[i]*")"),
      ylab=expression("Extinction Rate (r"[i]*")"),
      pe_cex=2)
title(main="C.",adj=0,family="serif",font=2)
peplot(mod3,"a10slsr",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
      pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
      ci_lty=2,xlab=expression("Continental Weathering Rates (r"[i]*")"),
      ylab=expression("Extinction Rate (r"[i]*")"),
      pe_cex=2)
title(main="D.",adj=0,family="serif",font=2)
peplot(mod3,"a10sdsl",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
      pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
      ci_lty=2,xlab=expression("Sea Level Instability (r"[i]*")"),
      ylab=expression("Extinction Rate (r"[i]*")"),
      pe_cex=2)
title(main="E.",adj=0,family="serif",font=2)

par(mfrow=c(1,1))
peplot.l(mod1,"a2sdc",plot_points="p",ylim=NULL,pch_col=ext_int_color_colors,
        pch_ty=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
        ci_lty=2,xlab=expression("Carbon Cycle Instability (r"[i]*")"),
```

```
          ylab=expression("Extinction Rate (r"[i]*")"),
          pe_cex=2)
    title(main="C.",adj=0,family="serif",font=2)

#within-bin sd carbon and sea level boxplots and ANOVA

  me<-c(0,0,1,0,0,0,1,1,0,0,1,0,0,1,0,0)
  dat_me<-cbind(me,wib.sd.carb[2:6])
  dat_me<-cbind((rownames(dat_me)),dat_me)
  colnames(dat_me)<-c("extint","me","0-2","2-4","4-6","6-8","8-10")
  dat_me$extint<- as.factor(dat_me$extint)

  dat_me$extint <- factor(dat_me$extint,
                       levels = c("Late Pliocene","Eocene-Oligocene","End-Cretaceous",
                                  "Cenomanian-Turonian","Tithonian","Toarcian","End-Triassic",
                                  "End-Permian","Guadalupian","Early-Serpukhovian","End-
Devonian",
                                  "Givetian-Frasnian","Eifelian-Givetian","End-Ordovician",
                                  "Series 3-Furogian","Stage 3-Stage 4"))
  df=melt((dat_me),id.vars=c("extint","me"),na.rm=FALSE)

  dat_me2<-cbind(me,wib.sd.sl[2:6])
  dat_me2<-cbind((rownames(dat_me2)),dat_me2)
  colnames(dat_me2)<-c("extint","me","0-2","2-4","4-6","6-8","8-10")
  dat_me2$extint<- as.factor(dat_me2$extint)

  dat_me2$extint <- factor(dat_me2$extint,
                        levels = c("Late Pliocene","Eocene-Oligocene","End-Cretaceous",
                                   "Cenomanian-Turonian","Tithonian","Toarcian","End-Triassic",
                                   "End-Permian","Guadalupian","Early-Serpukhovian","End-
Devonian",
                                   "Givetian-Frasnian","Eifelian-Givetian","End-Ordovician",
                                   "Series 3-Furogian","Stage 3-Stage 4"))
  df2=melt((dat_me2),id.vars=c("extint","me"),na.rm=FALSE)

  boxplot(df$value ~ df$variable,xlab="2 Myr Bins From 0 to 10 Myrs",
        ylab="Carbon Cycle Instability")

  # Add data points
  mylevels <- levels(df$variable)
  levelProportions <- summary(df$variable)/nrow(df)
  for(i in 1:length(mylevels)){

    thislevel <- mylevels[i]
    thisvalues <- df[df$variable==thislevel, "value"]
```

```
    # take the x-axis indices and add a jitter, proportional to the N in each level
    myjitter <- jitter(rep(i, length(thisvalues)), amount=levelProportions[i]/2)
    points(myjitter, thisvalues, pch=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16), col=ext_int_color_colors)
    legend("topright",
          legend = d2$extint,
          col =ext_int_color_colors,
          pch = c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16),
          bty = "y",pt.cex = 0.79,cex = 0.29,text.col = "black",
          horiz = F ,inset =F, ncol=2,title="Extinction Interval")

  }
  title(main="A.",adj=0,family="serif",font=2)

  #Sea Level

  boxplot(df2$value ~ df2$variable,xlab="2 Myr Bins From 0 to 10 Myrs",
         ylab="Sea Level Instability")

  # Add data points
  mylevels <- levels(df2$variable)
  levelProportions <- summary(df$variable)/nrow(df2)
  for(i in 1:length(mylevels)){

    thislevel <- mylevels[i]
    thisvalues <- df2[df2$variable==thislevel, "value"]

    # take the x-axis indices and add a jitter, proportional to the N in each level
    myjitter <- jitter(rep(i, length(thisvalues)), amount=levelProportions[i]/2)
    points(myjitter, thisvalues, pch=c(16,16,0,16,16,16,1,2,16,16,5,16,16,6,16,16), col=ext_int_color_colors)
  }
  title(main="B.",adj=0,family="serif",font=2)

#ANOVA
carb.short <-wib.sd.carb[,-1]
colnames(carb.short)<-c("two","four","six","eight","ten")
carb.long <- melt(carb.short, variable.name = "interval", value.name = "sdcarb")
mod4<-aov(sdcarb~interval,data=carb.long)
summary(mod4)
pairwise.t.test(carb.long$sdcarb,carb.long$interval,pool.sd=TRUE,p.adjust.method="bonferroni")

sl.short <-wib.sd.sl[,-1]
colnames(sl.short)<-c("two","four","six","eight","ten")
sl.long <- melt(sl.short, variable.name = "interval", value.name = "sdsl")
```

```
mod5<-aov(sdsl~interval,data=sl.long)
summary(mod5)
pairwise.t.test(sl.long$sdsl,sl.long$interval,pool.sd=TRUE,p.adjust.method="bonferroni")

###Future estimates

    #new df
    dat1.2<-subset(dat1,orgage<=2.4)
    rownames(dat1.2)<-NULL

    #Last 2Myrs
    #subset time and abiotic proxies
    myr2orgage<-(-1*dat1.2$orgage[subset=dat1.2$orgage>=0&dat1.2$orgage<=2])
    myr2dox<-dat1.2$dox[subset=dat1.2$orgage>=0&dat1.2$orgage<=2]
    myr2dcarb<-dat1.2$dcarb[subset=dat1.2$orgage>=0&dat1.2$orgage<=2]
    myr2stron<-dat1.2$stron[subset=dat1.2$orgage>=0&dat1.2$orgage<=2]
    myr2sl<-dat1.2$sl[subset=dat1.2$orgage>=0&dat1.2$orgage<=2]
    myr2patc<-dat1.2$patc[subset=dat1.2$orgage>=0&dat1.2$orgage<=2]

    #plots, what's happening in the bins?
    plot(myr2orgage,myr2dox)
    plot(myr2orgage,myr2dcarb)
    plot(myr2orgage,myr2stron)
    plot(myr2orgage,myr2sl)
    plot(myr2orgage,myr2patc)

    mmo<-mean(myr2dox,na.rm=TRUE)
    msdmo<-sd(myr2dox,na.rm=TRUE)
    msdc<-sd(myr2dcarb,na.rm=TRUE)
    lm1sr<-lm(dat1.2$orgage~dat1.2$stron,na.action=na.omit)
    mslsr<-((lm1sr$coefficients[2]))
    msdsl<-sd(myr2sl,na.rm=TRUE)
    #mmc<-mean(dat1.2$patc[subset=dat1.2$orgage>=0&dat1.2$orgage<=2],na.rm=TRUE)
    a2mbt<-data.frame(a2mo=mmo,a2sdo=msdmo,a2sdc=msdc,a2slsr=mslsr,a2sdsl=msdsl)
    row.names(a2mbt)<-"ext.now"
    ndf<-dat.alr2[1:6]
    modm<-lm(extrate~.,data=ndf[,c(1,4,6)])
    ext_now<-predict(modm,newdata=a2mbt[c(3,5)],interval="predict")
    ext_now

#corr between extinction magnitude and number of data points
#corr dataframe
sapply(cpaa, class)
is.numeric(dat.alr2$extrate)
extmag<-(dat.alr2$extrate)
cordf<-cbind(extmag,cpaa[,c(1,4,7,10)])
```

```
sapply(cordf, class)
i <- c(1:5)
cordf[ , i] <- apply(cordf[ , i], 2,            # Specify own function within apply
                    function(x) as.numeric(as.character(x)))
sapply(cordf, class)

#normality
describe(cordf)
shapiro.test(cordf$extmag) #almost sig
shapiro.test(cordf$a2ox) #sig
shapiro.test(cordf$a2carb) #sig
shapiro.test(cordf$a2stron) #not sig
shapiro.test(cordf$a2sl) #not sig

p <- ggpairs(cordf,
             upper = list(
               continuous = wrap('cor', method = "spearman")
             ),
title="Scatter Plot, Correlation, and Distribution Matrix")
print(p)

#spearman rank correlation analyses (data above are not normally distributed)
cor.test(cordf$extmag,cordf$a2ox,method="spearman",exact=TRUE)
cor.test(cordf$extmag,cordf$a2carb,method="spearman",exact=TRUE)
cor.test(cordf$extmag,cordf$a2stron,method="spearman",exact=TRUE)
cor.test(cordf$extmag,cordf$a2sl,method="spearman",exact=TRUE)
```