Exploring alternative methods for reconstructing temperatures using the clumped isotopes paleothermometer

Quan Gan
qgan@email.arizona.edu

School of Information, University of Arizona, USA

May 2022

Abstract

In the geosciences, 'clumped isotopes' paleothermometer (Δ_{47}) is widely used for inferring temperature reconstructions. Most of previous studies have used ordinary least squares linear models or York regressions to explore the relationship between Δ_{47} and Temperature. Recently, Román Palacios et al. [1] proposed the first Bayesian linear model for calibrating the Δ_{47} paleothermometer and reconstructing temperatures based on known Δ_{47} values. In this study, the proposed Bayesian models were shown to outperform other classic linear regression models, mainly in terms of accuracy and precision during temperature reconstruction. In this document, we use the models from Román Palacios et al. [1] as a baseline while exploring additional linear regression models that could potentially improve the basic toolbox for inferring temperature reconstructions based on the 'clumped isotopes' paleothermometer. Eventually, we compare the explored linear regression models to the baseline models in Román Palacios et al. [1]. We found the Quantile Regression and Robust Regression can significantly improve the accurary and precision during temperature reconstruction from synthetic datasets.

Keywords: Temperature Reconstruction, Bayesian Linear Regression, Ordinary Linear Regression, Error-in-variables Models

1 Introduction

Records of past temperatures are generally inferred using geochemical proxies. By performing temperature reconstructions, researchers can develop many empirical studies about climate change and subsequently construct different models to help us understand the climate variability and project future climate change. Most research [2–7] implementing regression models to calibrate paleoclimatic proxies have used either ordinary least squares linear regression or error-in-variables models. In fact, these models are used mathematically model the relationships between temperature and proxies.The measurement errors in the variables affect researchers to discover the correct relationship so increasing precision and accuracy of temperature reconstruction is the principal goal of related studies.

Clumped isotopes [8] is a paleotemperature proxy based on measurements of degree of ordering of ^{13}C and ^{18}O into bonds with each other in lattices of carbonate minerals, which is Δ_{47}. Clumped isotopes thermometry [1, 5, 8] relies only on thermodynamics to determine both the carbonate formation temperature and the oxygen isotopic composition of source water from a single carbonate sample measurement. Furthermore, unlike more traditional carbonate-based proxies, the clumped isotope paleothermometer does not rely on assumptions about the phase or the $\delta^{18}O_water$. The temperature dependence of carbonate clumped isotope thermometry has led to applications as broad-ranging as evolutionary biology, paleoclimate, and paleoaltimetry. In addition, it is suitable for interpolation and even modest extrapolation.

Machine learning techniques could improve the existing models of relationship between Clumped isotopes Δ_{47} and temperature T. Recent research have found that Δ_{47} scaled linearly with $1/T^2$ across temperature ranges of $0 - 100^oC$, $\Delta_{47} = \alpha + \beta * Temperature$. This means that one could deploy different linear regression model to calibrate the temperature based on Δ_{47} values. Machine learning algorithms can help us detect the outliers in the data sets and generate reliable synthetic data to validate the existing models. In addition, some machine learning techniques could also enhance existing models by considering the measurement errors among data points.

In this project, we aim to validate model performance in the study of Román Palacios et al. [1]. Furthermore, we explore additional models such as Orthogonal Distance Regression [9], Least Squares Monte Carlo with linear regression [10], Quantile Regression [11], Robust Regression [12], and Theil-sen Regression [13] as a reference in the clumped isotopes field. First, we fitted the previous classical models that include an ordinary least squares linear regression, York model, Deming regression, and Bayesian linear regression as the baseline models from BayClump [1]. Second, we explored additional linear regression models that might also contribute to improve accuracy and precision during parameter estimation based on 'clumped isotope' data. The results demonstrate the Quantile Regression, Robust Regression, and Theil-sen Regression greately improve the accuracy and precision of temperature reconstruction. All works and codes have been pushed in a Github repository, `https://github.com/qgan7125/Capstone_temperature_reconstructions.git`.

1.1 Related Work

Multiple studies have applied plenty linear regression models to calibrate the temperature based on Δ_{47} [1–7, 14]. We could classify these linear regression models to three categories. Each of these approaches have their own advantages and limitations. We discuss the models' framework in detail in the following sections.

Classical Linear Regression Models: The most commonly used linear regression model is the ordinary least squares linear regression. This model has been widely implemented in the all kinds of statistical packages. The ordinary least squares linear regression could help us find the coefficient α and β_i parameters of multiples variables in linear relationship to estimate the relationship. However, the ordinary least squares linear regression is sensitive to the outliers and could not take account for error in $10^6/T^2$ despite the facet that both errors in the clumped isotope and temperature measurements are used for deriving

Variable	Low Level	Intermediate Level	High Level
Δ_{47} across labs	0.0125%	0.0225%	0.0275%
Δ_{47} by instrument error	0.0025%	0.0075%	0.0125%
$1/T^2$	0.25^o C	2^o C	5^o C

Table 1: Δ_{47} and $1/T^2$ error levels

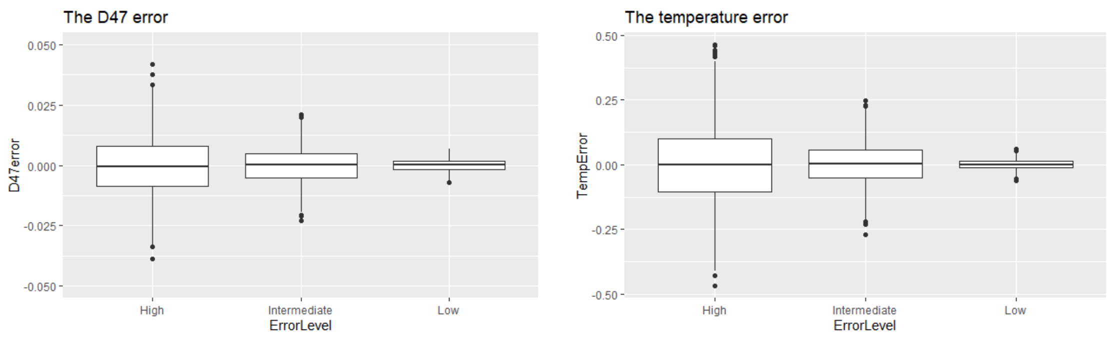
calibrations. In the clumped isotopes field, the magnitude of uncertainty in both variables, Δ_{47} and T varies for different calibration datasets.

Error-in-variables Models: Error-in-variables models is a regression model that account for the measurement errors in the variables. The common error-in-variables models are York regression and Deming regression. The original idea is from the 1979 paper [15] based on maximum likelihood estimation (MLE) of regression parameters. More rencently, York et al.[16] proposed the York regression. Deming regression [17, 18] is similar to the York model. Both the Deming and York models account for the measurement errors in the both variables. However, their performance in clumped isotopte datasets still need to be examined.

Bayesian Linear Regression Models: In the study of Román Palacios et al. [1], the authors proposed three different Bayesian Linear Regression Models. These models included a Bayesian simple Linear model, Bayesian linear with errors, and a Bayesian linear mixed with errors. These models were shown to improve the accuracy and precision in reconstructing temperatures under the clumped isotopes paleothermometer. The research shows these Bayesian models outperformed the transitional linear regression model in the clumped isotopes field. These models could improve their accuracy by using prior from previous models. However, based on the framework of Bayesian method, these Bayesian models have their limitations. In some conditions, Bayesian linear regression models might not reconstruct the correct temperature by posterior prediction.

2 Data

In this study, the data are three simulated datasets that come from the study of Román Palacios et al. [1]. The three simulated datasets are not the original datasets from Román Palacios study but were simulated based on the assumption in their study. Three simulated datasets are assumed as different scenarios, low, intermediate, and high levels of error in Δ_{47} and temperature. The true coefficient of the datasets are 0.0369 for the slope (β) and 0.268 for the intercept (α). In each dataset, there are 1000 data points for six features: Δ_{47} true value, $1/T^2$ true value, simulated Δ_{47}, simulated $1/T^2$ and their uncertainties. The table 1 contains the measurement errors in both variables of the simulated datasets. The distribution of Δ_{47} and temperature along with their associated measurement errors are displayed in the figure 1. The datasets provided by professor Cristian will not be disclosed and not be used for other purposes without his permission. All data are stored in the Data folder of GitHub repository. ”Dataset_S1_Mar8.csv” file contains the low error dataset. ”Dataset_S2_Apr20_V2.csv” file contains the intermediate error dataset. ”Dataset_S3_Apr20_V2.csv” file contains the high error dataset.



(a) (b)

Figure 1: (a) The Δ_{47} error distributions across the different datasets. (b) The temperature error distribution across the different datasets.

3 Methods and Models

All models were fitted with R version 4.1.1 [19]. Below are the relevant packages used for the analyses. For non-Bayesian models, uncertainty in regression parameters was estimated using the bootstrapping methods by applying R *boot* package [20] with 1,000 replicates with replace set to *TRUE*. For each replicate, data was splitted to 70% as train set 30% as test set. We calculated the train mean square error and test mean square error. The results (α, β, train MSE, and test MSE) of each fitted model were the medians of all replicates with 95% confidence intervals. All works and codes had been pushed in a private GitHub repository. In addition, we decided to contribute to the BayClump project [1]. BayClump is a self-contained R shiny Dashboard application. The new explored model might be merged to the BayClump application.

3.1 Baseline Models

All code and results for baseline models could be found in the Baseline_models folder in the GitHub repository. `https://github.com/qgan7125/Capstone_temperature_reconstructions/tree/main/Baseline_models`.

Ordinary Least Squares Linear Model: The first baseline model by applying *lm* function in the stats R package version 4.1.1 [19] with default parameters. This model fits linear regression with synthetic observed variables and do not specify the true variables and their measurement errors. This model aims to minimize the sum of the squares of the differences between Δ_{47} and $10^6/T^2$ by assuming a linear relationship.

York Regression: Second, we fitted a York regression model by *york* function provided by IsoplotR R package version 4.2 [21]. The *york* function is based on the unified regression algorithm of york [16]. The approach takes account of the measurement errors in both Δ_{47} and $10^6/T^2$ which vary across each data point and are weighted by correlation coefficient of Δ_{47} and $10^6/T^2$. Although this model is based least squares, it yields results that are consistent with maximum likelihood estimates as shown in Titterington and Halliday [15]. In this project, we only simulated Δ_{47} and $10^6/T^2$ and their measurement errors to fit the model.

Deming Regression: Third, we fitted a Deming regression model implemented by the *deming* function in deming R package version 1.4 [17]. This regression model is similar to the york model to find the maximum likelihood estimates line relating Δ_{47} and $10^6/T^2$ with their measurement error by minimize the sum of squared residuals but deming regression uses the constant ratio of variance in Δ_{47} and $10^6/T^2$ across each data point. Comparing to the parameters to fit york regression, deming regression uses the constant standard deviation of uncertainties and observed variables to fit the models.

***Bayesian linear with errors*:** Fourth, we fitted a Bayesian linear with errors model. Although Román Palacios et al. [1] provided three Bayesian models, which have achieved good performances on the simulated datasets, we only select one of them as the baseline model. The differences between Ordinary Least Squares Linear Model and Bayesian linear with error are to consider the measurement errors in the variables and to combine of coefficient prior distribution to fit the model. The prior distribution come from Román Palacios et al. and previous related search [1]. The mathematical definition of the model is below.

$$
\begin{aligned}
\Delta 47 &\sim Normal({\Delta_{47}^{true}}_i, \sigma_{47i}^2) \\
{\Delta_{47}^{true}}_i &\sim Normal(\mu_i, \sigma^2) \\
\mu_i &= \alpha + \beta \frac{10^6}{T^2}_i \\
\frac{10^6}{T^2}_i &\sim Normal(\frac{10^6}{T^2}_i^{true}, \sigma^2_{\frac{10^6}{T^2}_{true} i}) \\
\frac{10^6}{T^2}_i^{true} &\sim Normal(0, 10^{-3} \\
\alpha &\sim Normal(0.231, 0.065) \\
\beta &\sim Normal(0.039, 0.004) \\
\sigma^2 &\ Gamma(10^{-3}, 10^{-3}) \\
i &= 1,..., \text{N, where i is an index of each sample}
\end{aligned}
$$

The Bayesian model is implemented using the rstan R package and stan [22], which use C++ library for Bayesian modeling and inference. In our Bayesian model, we set the number of iterations to 2000 and sample using four chains. In addition, we also explored *brm* function in brms R package [23–25] to fit the Bayesian linear model with errors. The brms R package also uses stan to fit the Bayesian models. When we didn't specify the measurement errors in the Bayesian linear model, brm function yielded the same results as we observed from rstan R package without account for the measurement errors. Unfortunately, when we tried to add measurement errors in both variables by the *me* function in the brms R package, we could not use *brm* to recover our desired results. The α and β extremely differed from true parameters.

3.2 New Explored Models

We also explored below models to analyze if they could also contribute to improve the precision and accuracy of temperature reconstruction.

Least Squares Monte Carlo with linear regression**:** Fifth, we fitted a least squares monte carlo model. Monte Carlo simulation [26] is a method to predict the probability of different outcomes when the intervention of random variables is present. This model is similar to the ordinary least regression linear model presented above with bootstrapping technique. The bootstrapping technique samples data from the original dataset with duplicates allowed. In the Least Squares Monte Carlo, we used *lm* function to fit the model but each data points of Δ_{47} and $10^6/T^2$ is sampled from true Δ_{47} and $10^6/T^2$ and their measurement error distribution rather than the original simulated Δ_{47} and $10^6/T^2$.

Orthogonal Distance Regression**:** Sixth, we explored the orthogonal distance regression implemented by odregress function in pracma R package version 2.3.8 [27]. Unlike calculating least square distance in the Ordinary Least Squares Linear Model. The orthogonal distance regression model [9] minimizes the sum of squared perpendicular distance both on the Δ_{47} and $10^6/T^2$. Orthogonal regression assumes that the true values of the dependent and independent variables are linearly related. A slight inaccuracy has been introduced to the observed values of Y and X. Orthogonal regression entails finding a line that minimizes the equation below given n pairs of observed values.

Quantile Regression**:** Seventh, Quantile Regression is a model to fit the linear regression by the quantile. We used the *qr* function in the quantreg version 5.88 R package [28]. The normal linear regression uses constant β coefficients to find the relationship. In the Quantile Regression mode, β coefficients become functions with a dependency on the quantile. Quantile regression minimizes a weighted sum of absolute residuals. The mathematical definition of the model is below.

$$Q_\varsigma(y_i) = \alpha(\varsigma) + \beta(\varsigma)x_i \tag{1}$$

Robust Regression**:** Eighth, we used the *rlm* function in the MASS version 7.3-56 R package [29]. The *rlm* function is inherited from *lm* linear regression but uses iterated reweighted least squares (IWILS) to fit the model. The iterated reweighted least squares is used to solve certain optimization problems with objective functions of the form of a p-norm.

Theil-sen Regression**:** Finally, we used *theilsen* function from the deming R package version 1.4 [17]. Theil-sen regression in the deming R package is also a robust regression model but is different from the Robust regression. From the details of the deming package, one way to characterize the slope of an ordinary least squares line is that $\rho(x, r) = 0$, where ρ is the correlation coefficient and r is the vector of residuals from the fitted line. Theil-sen regression replaces ρ with Kendall's τ, a non-parametric alternative. It is resistant to outliers while retaining good statistical efficiency.

4 Model Performances

we used different methods to demonstrate the model performances. The section 4.1 and 4.2 state the model performances on 1,000 data points. The section 4.3 demonstrates the model performances on the small dataset (50 data points).

Error level scenario	Median Intercept	Intercept Std Error	α 95% CI	Median Slope	Slope Std Error	β 95% CI
Low error	0.26389	4.24e-05	0.2638 0.2640	0.0372	3.4e-06	0.03718 0.0372
Intermediate error	0.2697	8.77e-05	0.2696 0.27	0.03678	7.1e-06	0.03677 0.0368
High error	0.2656	1.09e-04	0.2654 0.2658	0.03717	8.9e-06	0.03714 0.03718

Table 2: Least Squares Monte Carlo with linear regression results on 1000 data points

4.1 Baseline Model Performances with 1,000 data points

Across all baseline models, Ordinary Least Squares Linear Model, York Regression, Deming Regression, and Bayesian linear with errors, we could discover that not all models' confidence intervals (CI) overlap with the true parameters (α and β). For the high error level dataset, york regression and deming regression can not overlap with true parameters (α and β). York regression (α CI(0.1625 - 0.2219), β CI(0.0411-0.0459)), and Deming regression (α CI(0.1627 - 0.2219), β CI(0.0411-0.0458)) have very similar performances on the high error level dataset and both underestimated the intercept and overestimated the slope. In addition, Deming regression was least accurate and precise across all models. The Bayesian linear with errors model could not recover the coefficient for the low error level dataset, the intercept CI (0.263-0.265) was underestimated and the slope CI(0.0371 - 0.0373) was overestimated. Ordinary Least Squares Linear Model had the best performances and could recover the true parameters with most accurate and precise. The median train MSE and median test MSE were similar across all baseline models.

4.2 New Explored Model Performances with 1,000 data points

***Least Squares Monte Carlo with linear regression*:** This model could not recover the intercept α and slope β. The result is listed in table 2. For low error level and high error level dataset, Least Squares Monte Carlo with linear regression underestimated the intercept and overestimated the slope. For the intermediate error level, Least Squares Monte Carlo with linear regression overestimated the intercept and underestimate slope. The only advantage of Least Squares Monte Carlo with linear regression was that the lowest variances in both parameters' confidence interval. Therefore, Least Squares Monte Carlo with linear regression is not suitable for temperature reconstruction. The possible reason for this result is because Least Squares Monte Carlo with linear regression derived data points from true values with distribution of measurement errors rather than considering the actual measured values. In addition, in practice, when we measure Δ_{47} and $10^6/T^2$, we could not decide what is the accurate variances. Least Squares Monte Carlo with linear regression could only work for the simulated dataset.

***Orthogonal Distance Regression*:** In the low error level and high error level dataset, Orthogonal Distance Regression performed well as same as the Ordinary Least Squares Linear Model. The result is listed in the table 3. However, from the slope confidence interval (β 0.03705 - 0.03772), Orthogonal Distance Regression overestimated the slope in the intermediate error level dataset. Although Orthogonal Distance Regression overestimated the slope but it was still close to the true parameter 0.0369 so we still consider Orthogonal Distance Regression as a possible method for temperature reconstruction fields.

Error level scenario	Median Intercept	Intercept Std Error	α 95% CI	Median Slope	Slope Std Error	β 95% CI
Low error	0.264	7.13e-05	0.2612 0.2696	0.03719	5.70e-06	0.03675 0.03746
Intermediate error	0.2702	1.421e-04	0.2573 0.2692	0.03674	1.15e-05	0.03706 0.03772
High error	0.26353	1.63e-04	0.2533 0.2729	0.0373	1.33e-05	0.03645 0.0381

Table 3: Orthogonal Distance Regression results on 1000 data points

Error level scenario	Median Intercept	Intercept Std Error	α 95% CI	Median Slope	Slope Std Error	β 95% CI
Low error	0.2649	9.67e-05	0.2642 0.2743	0.03712	7.80e-06	0.0364 0.0372
Intermediate error	0.2731	2.088e-04	0.253 0.2721	0.0365	1.72e-05	0.03684 0.03806
High error	0.2699	1.998e-04	0.2612 0.2848	0.03684	1.61e-05	0.03566 0.03764

Table 4: Quantile Regression results on 1000 data points

***Quantile Regression*:** Quantile regression could recover true parameters based on its confidence intervals. The result is listed in table 4. In addition, we found Quantile regression performed a little bit better than Ordinary Least Squares Linear Model in the low error level and high error level.

***Robust Regression*:** Robust Regression also outperformed the Ordinary Least Squares Linear Model but performed worse than Quantile regression. The result is listed in the table 5.

***Theil-sen Regression*:** Theil-sen Regression performed almost same as the Ordinary Least Squares Linear Model and also recover the true parameters. The result is listed in the table 6.

By comparing the baseline models and new explored models, we found new explored models except Least Squares Monte Carlo with linear regression could contribute to improve the accuracy and precision of temperature reconstruction. And we recommend the researchers could attempt to use these models to recover the relationship between Δ_{47} and $10^6/T^2$. In addition, no models were not consistently better than other models on different

Error level scenario	Median Intercept	Intercept Std Error	α 95% CI	Median Slope	Slope Std Error	β 95% CI
Low error	0.2645	7.2e-05	0.2623 0.2708	0.03716	5.70e-06	0.03668 0.03737
Intermediate error	0.2631	1.486e-04	0.2577 0.274	0.03724	1.2e-05	0.03635 0.03774
High error	0.2612	1.617e-04	0.2544 0.2749	0.0374	1.32e-05	0.03625 0.03799

Table 5: Robust Regression results on 1000 data points

Error level scenario	Median Intercept	Intercept Std Error	α 95% CI	Median Slope	Slope Std Error	β 95% CI
Low error	0.264	7.46e-05	0.2626 0.2712	0.0372	5.90e-06	0.03659 0.03738
Intermediate error	0.2718	1.445e-04	0.258 0.2703	0.03663	1.16e-05	0.037 0.03763
High error	0.2654	1.77e-04	0.2558 0.2773	0.03719	1.41e-05	0.03625 0.038

Table 6: Theil-sen Regression results on 1000 data points

error level. We need to choose suitable models for different error scenarios.

4.3 Model Performance on 50 Data Points

To better understand the model performances, we sampled 50 data points from the low error synthetic dataset. $\Delta_{47} = 0.0025\%_o$ and $10^6/T^2 = 0.25^oC$. Then using the small synthetic data set to fit our nine models. In the figure 2, the white areas are the true parameter intercept and slope with measurement error. The yellow areas are the fitted model coefficient parameters. The black points are the sampled 50 data points.

A-D in the figure 2 are the baseline models, just like Román's Research [1], the Ordinary Least Squares Linear Model and Bayesian linear model with errors did outperform York regression and Deming regression models. The Ordinary Least Squares Linear Model and Bayesian linear model with errors model performed well in these 50 data points, which overestimated slope ∼3.5% and underestimated intercept ∼5.5%. The Ordinary Least Squares Linear Model overestimated slope ∼ 2.7% and underestimated intercept ∼4.4%. The York regression overestimated slope ∼6.5% and underestimated intercept ∼9.7%. The Deming regression model had the highest variances across all models to underestimate the slope ∼ 15% and to overestimate intercept ∼ 24%. In addition, Deming model took the longest time to train the model. If we observe the confidence intervals, only the results of Bayesian Linear Model with errors overlapped with true parameters.

E-I are the new explored models in this study. The Least Squares Monte Carlo with linear regression (F) outperformed than Bayesian linear model with errors but Least Squares Monte Carlo with linear regression still overestimated slope ∼3.5% and underestimated intercept ∼5.4%. The result is expected. Based on the framework of monte carlo, this model actually sampled data from the true independent and dependent variables. Therefore, the white line and yellow line are overlapped but the least squares monte carlo is not practical. We use synthetic datasets in this study so we know the true variables. When we use new measured variables, we actually have no idea about the measurement errors and true variables. The Orthogonal Distance Regression (E) has the same performance with the Ordinary Least Squares model, which both overestimated slope ∼ 2.7% and underestimated intercept ∼4.4%. The Theil-sen regression (I) and Robust Regression (H) performed well and slightly better than the Ordinary Least Squares model. They both overestimated slope ∼1.3% and underestimated intercept ∼1.9% -2.2%7. In the current temperature reconstructions research, only few researchers [30, 31] use robust regression to explore the relationship between proxy and temperature. The Quantile regression is the best model across all practical models, which overestimated slope ∼1% and underestimated intercept ∼1.6%. We recommend further studies to research how good the Quantile regression and Robust Regression are in clumped isotopes fields.

The another consideration is the whether the confidence intervals are overlapped with true parameters. Only Bayesian linear model with errors achieved it because the 50 data points model performances would be influenced by the method to sample data points from the original dataset. Therefore, the model performances on 50 data points are different with the results from the 1000 data points. We could conclude that the size of dataset would impact the model performances. In our study, to reproduce the result, we set seed as 3 and used the sample function with replace as FALSE to sample points.

5 Inverting Prediction from Δ_{47} to $10^6/T^2$

After we derived the calibration slopes and intercepts from the different regression models. We then reconstructed temperature by the coefficients and evaluated model performances by how accurate and precise the resulting temperature were close to the true value. In the inverting prediction, the temperature is unknown. We aimed to predict the previous independent variable temperature from the previous dependent variable Δ_{47}. We defined three Δ_{47} targets as $0.6\%_o$, $0.7\%_o$, and $0.8\%_o$. The true temperatures associated with these Δ_{47} targets are 60^oC, 19^oC, and -9.8^oC . For each Δ_{47} target scenario, we assumed the low, intermediate, and high error in them as $0.005\%_o$, $0.01\%_o$, and $0.02\%_o$. The temperature accuracy defined as the median distance to true temperature as ^oC. For example, we suppose the predicted temperature for $0.6\%_o$ Δ_{47} target plus error is 62^oC. The accuracy is the difference between true temperature the predicted temperature, $62^oC - 62^oC = 2^oC$. And the precision defined as the SE between Δ_{47} target and the Δ_{47} target with errors. For example, we first predicted temperature from Δ_{47} target $0.6\%_o$ and temperature from Δ_{47} target $0.6\%_o + 0.005\%_o$. The difference between these two temperatures are the SE for the models.

5.1 Inverting Prediction Framework

In this project, the models are classified as two categories as Bayesian models and non-Bayesian models so we had to use two different methods to do inverting prediction. The coefficient results (α and β) are special case. We could use both Bayesian methods and non-Bayesian methods to do inverting prediction. In order to achieve best accurate, we use the functions from BayClump project to calculate the coefficient and inverse prediction. For the new explored models, we followed the BayClump function syntax and adopt new R functions to train the models.

Bayesian method: Bayesian inversion prediction uses the updated posterior distribution from the original Bayesian linear model to predict temperature. We use *jags* function in the R2jags R package [32] to update the posterior distribution. The prior distribution defined below.

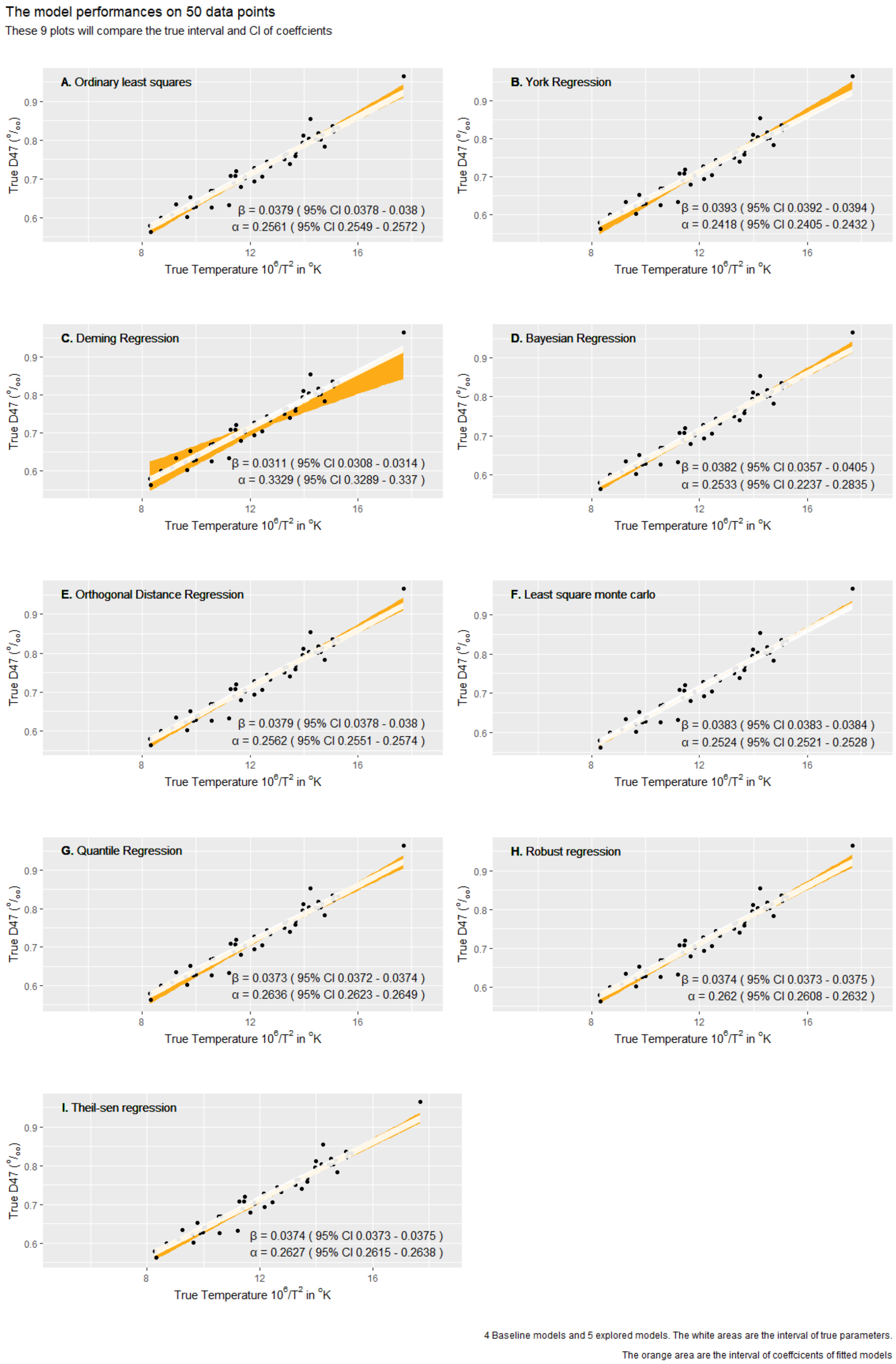


Figure 2: The 9 models' performances on 50 data points. The orange area are the confidence interval of the fitted model and the white area are the true coefficient parameters' confidence interval. The black points are the sampled 50 data points. The true slope = 0.0369 and intercept = 0.268

$$x_i \sim Normal(11, 0.394)$$
$$\Delta 47_i \sim Normal(\Delta 47_i^{predict}, \sigma)$$
$$Temperature_i = \sqrt{(\beta * 10^6)/(\Delta 47_i - \alpha)} - 273.15$$
$$Temperature_i^{predict} = \sqrt{Temperature_i}$$

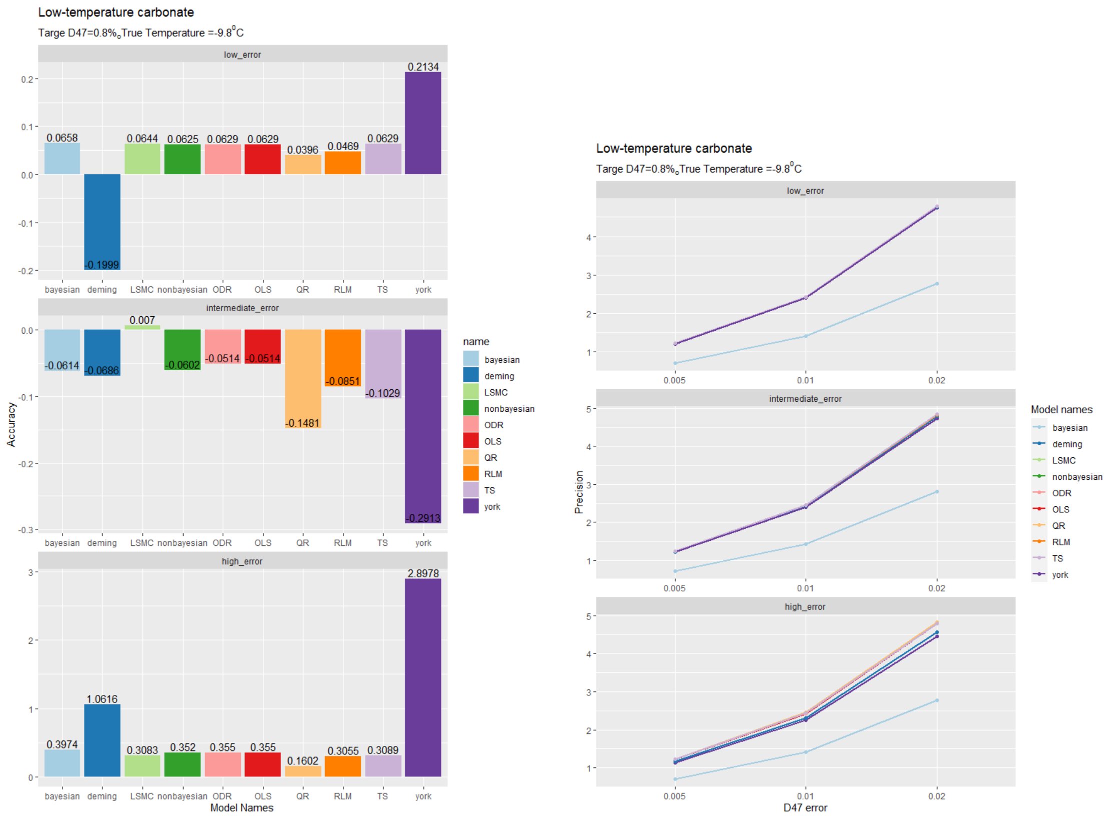
Non-Bayesian method: We used *nls* function in the stats R package [33] and *invest* function in the inverstr R package [34]. The *nls* function is to determine the nonlinear (weighted) least-squares estimates of the parameters of a nonlinear model. The *investr* function is to provide point and interval estimates for the unknown predictor value that corresponds to an observed value of the response.

5.2 Reconstruction for low temperature carbonates ($\Delta_{47} = 0.8\%_o$ and Temperature = -9.8^oC)

From the figure 3 (a) accuracy of nine models, Quantile regression and Robust Regression models outperformed all other models, which overestimated 0.0396^oC and 0.0469^oC in the low error scenario and overestimated 0.1602^oC and 0.3055^oC in the high error scenario. However, Quantile regression and Robust Regression had higher variances in the intermediate error scenario. The Least Squares Monte Carlo with linear regression performed best which overestimated 0.007^oC in the intermediate error scenario. Just like we observed above, Orthogonal Distance Regression and Least Squares Monte Carlo with linear regression and similar performances across all scenarios. The York Regression had least accurate across all scenarios. In the intermediate error scenario, there were more differences than other two scenarios. The Deming Regression became better. All models almost underestimated the predicted temperature. The new explored models, Quantile Regression, Robust Regression, and Theil-sen Regression display less accurate.

From the figure 3 (b) precision of nine models, the Bayesian reconstruction outperformed all other models across all error levels in the target Δ_{47}. The predicted temperature precision in the Deming regression and York Regression increased with the increased error level scenario in the train set. The Deming Regression performance was different with results in the Román's study [1].

From the result of accuracy and precision, Bayesian reconstruction was still the optimal model to reconstruct the temperature but we also suggest Quantile Regression and Robust Regression as the alternatives in the low temperature carbonates when Bayesian reconstruction cannot provide reliable coefficient results. We do not recommend the Least Squares Monte Carlo with linear regression because of it's impractical and can not provided reliable parameters in the 95% confidence interval, even though it performed very well on the intermediate error scenarios.



(a) (b)

Figure 3: (a) The median distances from predicted temperature to true temperature. (accuracy). (b) The standard error in the predicted temperature (precision). Temperature reconstruction performance across nine models. The coefficient derived from low error level dataset. The target Δ_{47} is $0.8\%_o$ with $0.005\%_o, 0.01\%_o, and 0.02\%_o$ errors and the true temperature is -9.8^oC

5.3 Reconstruction for intermediate temperature carbonates ($\Delta_{47} = 0.7\%_o$ and Temperature = 19^oC)

From figure 4 (b), we could observed the patterns as same as the precision in the figure 3 (b). The Bayesian reconstruction was still the most precise model.

From figure 4 (a), Quantile regression still outperformed all other models which underestimated 0.0698^oC in the low error level scenario. Bayesian linear model with errors and Robust regression performed also well, which underestimated 0.0794^oC and 0.0753^oC in the low error level scenario. Deming Regression had the least accurate which underestimated 0.43596^oC in the low error level scenario. York regression became more accurate in this intermediate temperature carbonates. The rest four models performed almost same. For the intermediate error level scenario, Deming regression became the most accurate model, which underestimated 0.1129^oC. York regression was the least accurate model, which also underestimated 0.67323^oC. The rest seven models all overestimated temperatures. It is noteworthy that the Quantile Regression was less accurate, which overestimated 0.2424^oC. In the high error scenario, Deming regression became the most inaccurate model, which underestimated 0.668^oC. The Quantile regression performed worse, which overestimated 0.3637^oC. Across all scenarios, the Ordinary Least Squares Linear Model and Orthogonal Distance Regression outperformed all other models.

In conclusion, we recommend using Bayesian linear models, Ordinary Least Squares Linear Model and Orthogonal Distance Regression reconstruct the temperature in the intermediate temperature carbonates.

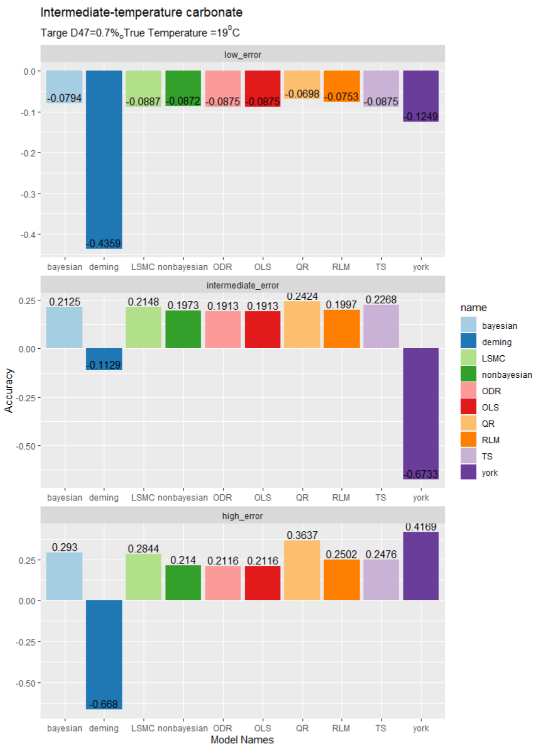
5.4 Reconstruction for high temperature carbonates ($\Delta_{47} = 0.6\%_o$ and Temperature = 60^oC)

In the high temperature carbonates, the precision of all models had no obvious differences. Unlike the accuracy results above, there was no pattern between accuracy and error level scenarios. In the low error and intermediate error scenarios, all models were less accuracy from -0.3539^oC to 0.9552^oC. In the high error scenario, all models except Deming regression, York regression, and Quantile Regression were highly accurate from -0.0885^oC to 0.1522^oC. Quantile Regression performed best in the low error scenario. The Deming regression performed best in the intermediate error scenario. And Bayesian linear model with errors, Robust Regression, and Theil-sen regression outperformed other models in the high error scenario. Therefore, we recommend Robust regression and Theil-sen regression as the alternate methods to reconstruct temperature in the high temperature carbonates.

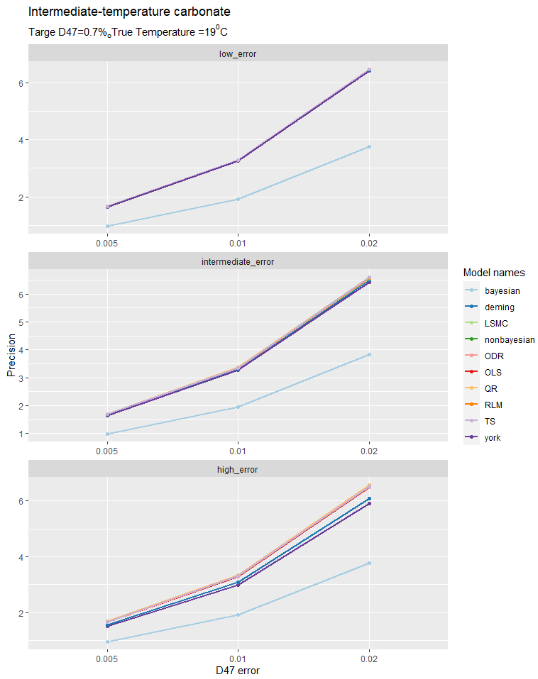
In the temperature prediction experiment, we compared the nine models performances in the different error level scenarios with different variances in the target Δ_{47}. Although different models' performances varied from the error levels and variances, the Bayesian linear model with errors was the best models in general. When Bayesian linear model with errors cannot recover the true parameters in the 95% confidence interval. We recommend to use Quantile Regression in the low error level and high error level scenarios and Robust Regression in the all scenarios.

6 Discussion

To review all results in the coefficient derivation and inverting prediction, we could

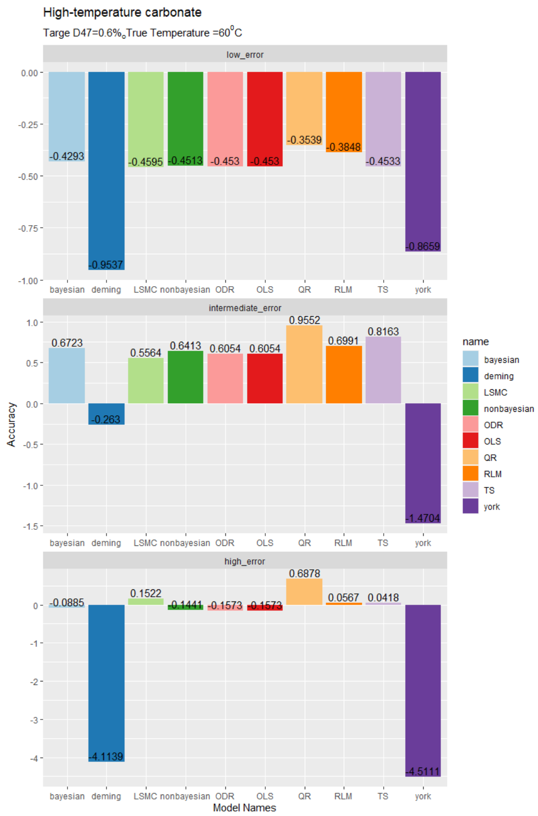


(a)

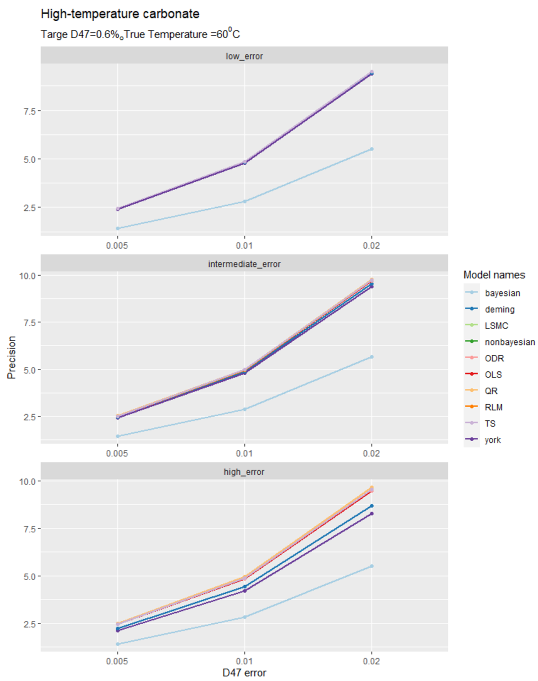


(b)

Figure 4: (a) The median distances from predicted temperature to true temperature. (accuracy). (b) The standard error in the predicted temperature (precision). Temperature reconstruction performance across nine models. The coefficient derived from low error level dataset. The target $\Delta 47$ is $0.7\%_o$ with $0.005\%_o, 0.01\%_o, and 0.02\%_o$ and the true temperature is 19^oC



(a)



(b)

Figure 5: (a) The median distances from predicted temperature to true temperature. (accuracy). (b) The standard error in the predicted temperature (precision). Temperature reconstruction performance across nine models. The coefficient derived from low error level dataset. The target $\Delta 47$ is $0.6\%_o$ with $0.005\%_o, 0.01\%_o, and 0.02\%_o$ and the true temperature is 60^oC

find the results yielded from intermediate error level scenario were relatively larger differed from the results from low error level scenario and high error level scenario. For the further research, we need to focus more on intermediate level scenario and intermediate temperature carbonates to improve the precision and accuracy. It is worth to generate new synthetic dataset and to redo every experiment. The project contributes to temperature reconstruction by finding new regression methods, Quantile regression, Robust Regression, and Theil-sen regression which could also improve the accuracy and precision in the clumped isotopes field. In the calibration step, Quantile Regression reduced uncertainty compared to York regression from 3.1%-17.72% to 0.74%-1.92% in α and from 1.12%-11.59% to 0.158%-1.05% in β. And the best performed models are Quantile Regression (low error scenario and high error scenario) and Bayesian linear model with error (intermediate error scenario). In the reconstruction steps, Quantile regression improved 27.93% accuracy and Robust regression improved 26.45% accuracy compared to York regression in low temperature carbonate with high error level train set. The theil-sen regression improved 7.45% accuracy compared to York regression in high temperature carbonate with high error level train set. The best performed models in the temperature reconstruction are Quantile regression (low error scenario and low error in Δ_{47}), theil-sen regression (high error scenario and high error in Δ_{47}), and Bayesian linear model with errors (all scenarios in combination with accuracy and precision).We also recommend the further research could explore these three regression method and validate their performances.

7 Acknowledgement

For each week meeting, Professor Cristian Roman Palacios provided wonderful feedback and good suggestions to make sure the project is heading in the right direction. I would like to express my special thanks of gratitude to my teacher, Cristian Roman Palacios (University of Arizona) for providing research guideline and writing support.

References

[1] C. Román Palacios, H. Carroll, A. Arnold, R. Flores, S. Petersen, K. McKinnon, and A. Tripati. "BayClump: Bayesian Calibration and Temperature Reconstructions for Clumped Isotope Thermometry". In: *Earth and Space Science Open Archive* (2021), p. 40. DOI: `10.1002/essoar.10507995.1`. URL: `https://doi.org/10.1002/essoar.10507995.1`.

[2] H Elderfield and G Ganssen. "Past temperature and δ18O of surface ocean waters inferred from foraminiferal Mg/Ca ratios". In: *Nature* 405.6785 (2000), pp. 442–445.

[3] B. Christiansen and F. C. Ljungqvist. "Challenges and perspectives for large-scale temperature reconstructions of the past two millennia". In: *Reviews of Geophysics* 55.1 (2017), pp. 40–96.

[4] E. A. Schauble, P. Ghosh, and J. M. Eiler. "Preferential formation of 13C–18O bonds in carbonate minerals, estimated using first-principles lattice dynamics". In: *Geochimica et Cosmochimica Acta* 70.10 (2006), pp. 2510–2529.

[5] P. Ghosh, J. Adkins, H. Affek, B. Balta, W. Guo, E. A. Schauble, D. Schrag, and J. M. Eiler. "13C–18O bonds in carbonate minerals: a new kind of paleothermometer". In: *Geochimica et Cosmochimica Acta* 70.6 (2006), pp. 1439–1456.

[6] R. A. Eagle, E. A. Schauble, A. K. Tripati, T. Tütken, R. C. Hulbert, and J. M. Eiler. "Body temperatures of modern and extinct vertebrates from 13C-18O bond abundances in bioapatite". In: *Proceedings of the National Academy of Sciences* 107.23 (2010), pp. 10377–10382.

[7] J. R. Kelson, K. W. Huntington, A. J. Schauer, C. Saenger, and A. R. Lechler. "Toward a universal carbonate clumped isotope calibration: Diverse synthesis and preparatory methods suggest a single temperature relationship". In: *Geochimica et Cosmochimica Acta* 197 (2017), pp. 104–131.

[8] J. M. Eiler. "Paleoclimate reconstruction using carbonate clumped isotope thermometry". In: *Quaternary Science Reviews* 30.25-26 (2011), pp. 3575–3588.

[9] P. T. Boggs and J. E. Rogers. "Orthogonal distance regression". In: *Contemporary Mathematics* 112 (1990), pp. 183–194.

[10] D. D. McCracken. "The monte carlo method". In: *Scientific American* 192.5 (1955), pp. 90–97.

[11] R. Koenker and G. Bassett Jr. "Regression quantiles". In: *Econometrica: journal of the Econometric Society* (1978), pp. 33–50.

[12] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust statistics: the approach based on influence functions*. Vol. 196. John Wiley & Sons, 2011.

[13] P. K. Sen. "Estimates of the regression coefficient based on Kendall's tau". In: *Journal of the American statistical association* 63.324 (1968), pp. 1379–1389.

[14] R. E. Came, J. M. Eiler, J. Veizer, K. Azmy, U. Brand, and C. R. Weidman. "Coupling of surface temperatures and atmospheric CO2 concentrations during the Palaeozoic era". In: *Nature* 449.7159 (2007), pp. 198–201.

[15] D. M. Titterington and A. N. Halliday. "On the fitting of parallel isochrons and the method of maximum likelihood". In: *Chemical Geology* 26.3-4 (1979), pp. 183–195.

[16] D. York, N. M. Evensen, M. L. Martınez, and J. De Basabe Delgado. "Unified equations for the slope, intercept, and standard errors of the best straight line". In: *American journal of physics* 72.3 (2004), pp. 367–375.

[17] T. Therneau. *deming: Deming, Theil-Sen, Passing-Bablock and Total Least Squares Regression*. R package version 1.4. 2018. URL: `https://CRAN.R-project.org/package=deming`.

[18] K. Linnet. "Estimation of the linear relationship between the measurements of two methods with proportional errors". In: *Statistics in Medicine* 9.12 (1990), pp. 1463–1473.

[19] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: `https://www.R-project.org/`.

[20] A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Applications*. ISBN 0-521-57391-2. Cambridge: Cambridge University Press, 1997. URL: `http://statwww.epfl.ch/davison/BMA/`.

[21] P. Vermeesch. “IsoplotR: a free and open toolbox for geochronology”. In: *Geoscience Frontiers* 9 (2018), pp. 1479–1493. URL: `https://doi.org/10.1016/j.gsf.2018.04.001`.

[22] Stan Development Team. *RStan: the R interface to Stan*. R package version 2.21.3. 2021. URL: `https://mc-stan.org/`.

[23] P.-C. Bürkner. “brms: An R Package for Bayesian Multilevel Models Using Stan”. In: *Journal of Statistical Software* 80.1 (2017), pp. 1–28. DOI: `10.18637/jss.v080.i01`.

[24] P.-C. Bürkner. “Advanced Bayesian Multilevel Modeling with the R Package brms”. In: *The R Journal* 10.1 (2018), pp. 395–411. DOI: `10.32614/RJ-2018-017`.

[25] P.-C. Bürkner. “Bayesian Item Response Modeling in R with brms and Stan”. In: *Journal of Statistical Software* 100.5 (2021), pp. 1–54. DOI: `10.18637/jss.v100.i05`.

[26] S. Nadarajah, F. Margot, and N. Secomandi. “Comparison of least squares Monte Carlo methods with applications to energy real options”. In: *European Journal of Operational Research* 256.1 (2017), pp. 196–204.

[27] H. W. Borchers. *pracma: Practical Numerical Math Functions*. R package version 2.3.8. 2022. URL: `https://CRAN.R-project.org/package=pracma`.

[28] R. Koenker. *quantreg: Quantile Regression*. R package version 5.88. 2022. URL: `https://CRAN.R-project.org/package=quantreg`.

[29] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York: Springer, 2002. URL: `https://www.stats.ox.ac.uk/pub/MASS4/`.

[30] C. Zeng, H. Shen, M. Zhong, L. Zhang, and P. Wu. “Reconstructing MODIS LST based on multitemporal classification and robust regression”. In: *IEEE Geoscience and Remote Sensing Letters* 12.3 (2014), pp. 512–516.

[31] C. Waelbroeck, L. Labeyrie, E. Michel, J.-C. Duplessy, J. F. Mcmanus, K. Lambeck, E Balbon, and M. Labracherie. “Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records”. In: *Quaternary science reviews* 21.1-3 (2002), pp. 295–305.

[32] Y.-S. Su and M. Yajima. *R2jags: Using R to Run ’JAGS’*. R package version 0.7-1. 2021. URL: `https://CRAN.R-project.org/package=R2jags`.

[33] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: `https://www.R-project.org/`.

[34] B. M. Greenwell and C. M. S. Kabban. “investr: An R Package for Inverse Estimation”. In: *The R Journal* 6.1 (2014), pp. 90–100. URL: `https://doi.org/10.32614/RJ-2014-009`.