Title
THE ROLE OF ENVIRONMENT IN THE EVOLUTION OF CRANIAL CAPACITY IN HUMANS

Author
Harnett, Joshua J

Publication Date
2024-07-24

THE ROLE OF ENVIRONMENT IN THE EVOLUTION OF CRANIAL CAPACITY IN HUMANS

By

Joshua Jeon Seon Harnett

A capstone project submitted for Graduation with University Honors

May 9, 2024

University Honors
University of California, Riverside

APPROVED

Dr. Sang-Hee Lee
Department of Anthropology

Dr. Richard Cardullo, Howard H Hays Jr. Chair
University Honors

ABSTRACT

Over the last 2 million years, the human brain has undergone an extraordinary evolutionary journey, expanding in size to nearly thrice that of our ancient ancestors, Homo habilis. This remarkable growth far surpasses the evolutionary changes seen in other primates, indicating a unique path of human development. Scientists have suggested a wide range of factors to explain the dramatic increase in brain size, including environmental, dietary, social, and climatic influences. This paper uses environmental data joined with hominid fossil data to create a prediction model using linear regression. Four different models are created, with the most significant model showing that environment variables along with time explain 21 percent of the variance found in cranial capacity over the past 7 million years.

ACKNOWLEDGEMENTS

I am deeply grateful to Dr. Sang-hee Lee, whose guidance has been indispensable for my research. Her commitment and passion for anthropological research are truly inspiring and have played a crucial role in our understanding of human evolution. I would also like to thank my community college professor, Dr. Scott Suarez, for sparking my interest in anthropology research. Additionally, I am thankful to the students and faculty of BCOE for the data science skills they have shared, which I will treasure and utilize throughout my career.

TABLE OF CONTENTS

INTRODUCTION

In the vast expanse of Earth's history, the environmental transformations brought about by geological events have played a pivotal role in shaping the course of human evolution. The collision of continents, the rise of majestic mountain ranges like the Himalayas, and the consequent shifts in global climate have been instrumental in altering the landscapes our ancestors once roamed. These monumental changes, such as the transformation of dense rainforests into vast savannas in Africa, set the stage for a series of evolutionary challenges and opportunities (Mercader 2002). This foundation is crucial for understanding the development of human cranial capacity and brain size, highlighting the importance of environmental factors in molding the habitats and ultimately the evolutionary paths of our ancestors.

The study of the evolution of human cranial capacity has been a topic for research for centuries. The various studies that have been conducted has shown that cranial development is not simple, and is subject to a multitude of possibilities (Hanken 1993; Willmore 2006; Chaline 2003). Some implications from what fossil records are available unveil some hints to us. One being how dietary shifts may have been a catalyst in shaping larger brains, as shown by the first genus Homo having dental reduction (Neubauer 2012). Another study shows how cranial capacity evolution in humans is a result of a dual inheritance system of biology and culture (Weber 2023). Kenneth Beals et al. (1984) suggests colder climates led to changes in cranial volume, imposing that climate influences cranial evolution. From these studies we may conclude that human cranial capacity evolution is complex with many contributing factors.

A great leap in human evolution came with the introduction of bipedalism, which greatly affected hominin locomotion (Sylvester 2006). This was a branching moment in our evolutionary tree where hominids adapted for greater efficiency for movement on the ground over climbing in trees (Thorpe 2007). How does this relate to cranial capacitance? A consequence of bipedalism is a reduction in pelvis width, which ultimately limits the cranial growth in newborns (Neubauer 2012). The relationship between the pelvic inlet and neonatal cranial size are very closely correlated (Rosenberg 2003). To answer the origin of bipedalism, a common theory is that Africa experienced climate shifts during Pleistocene and Pliocene epochs. This variability resulted in a mosaic of ecological settings, ranging from wet and lush environments to arid and open landscapes (Potts 2007). Hominids took a divergent path which led to higher efficiency for land travel. This is one example of the many evolutionary steps impacted by environments our ancient ancestors were subjected to. This example shows how the relationship between environmental factors and cranial capacity is complex, and involves understanding the mechanics of evolutionary adaptations, as well as the constraints they impose on anatomical development.

In examining the intricate relationship between climate, temperature, and human evolution, Ash (2007) provides compelling evidence that environmental factors significantly influenced the development of larger brain sizes in early humans. Ash's research suggests that 52% of the variation in skull size can be attributed to temperature variation, while 22% is explained by the distance from the equator. The underlying hypothesis suggests that the exposure to multiple habitats and the necessity to adapt to changing environments fostered the growth of certain alleles that increase brain size. These findings indicate a unique evolutionary path for humans, where the challenges posed by diverse and fluctuating climates were met not just with physical adaptations but through an unprecedented expansion of intellectual capabilities. Such a pivot toward solving problems through intellectual strategies marks a significant departure from the evolutionary strategies observed in other species and highlights the crucial role of environmental pressures in shaping human evolution.

This study aims to enhance understanding of human evolution by examining the role of environmental pressures in the development of larger brain sizes. By applying a machine learning approach, it analyzes how environmental factors have influenced evolutionary changes in human cranial capacity. The research suggests that the evolution of brain size was not a complex, multifaceted process but rather a linear one, primarily driven by adaptive responses to a dynamically changing environment. This unique evolutionary trajectory emphasizes a significant link between the environmental history of the planet and the development of human intellectual capabilities. Utilizing data regression analysis, this paper provides a comprehensive explanation of the patterns observed in human cranial capacity evolution, highlighting the pivotal role of environmental conditions. All the material used for this study can be found here Link to Google Drive.

METHODS

Researchers have made several assumptions about how environmental variables affect evolution in hominids. Some of these assumptions are how correlations between climatic events and evolutionary changes imply a causal relationship (Potts 2007; Ash 2007). Another being cognitive demands imposed by changing environments led to the selection of traits that improved survival, such as enhanced intellectual abilities. This involves an assumption that intellectual and behavioral flexibility was crucial for adapting to new challenges presented by environmental variability (Ash 2007).

A large portion of research in cranial capacity evolution focuses on environmental determinism, which presumes how climate variability was responsible for unique hominin traits such as bipedality. Recent theories and scientific discussions suggest that rapid climatic changes were crucial in driving increases in human brain size. These climatic changes are hypothesized as having shaped physical adaptations and cultural innovations among early human populations (Livingstone 2012). The resulting research suggests a strong link between climatic variability and the evolutionary success of early human populations. This comes as no surprise as evolution is driven by adaptations towards one's environment. In the case of dramatic climate variation, how does a species adapt accordingly, and how fast can it adapt?

In a study conducted by Craig Stockwell, he discusses how rapid evolutionary changes can occur within a few generations, and scales relevant to environmental pressures. Many variables of the environment play a part in how a species may adapt. Carbon dioxide levels may impact plant photosynthesis rates, which affects food availability (Stockwell 2003). In Jessica Ash’s study (2007), cooling trends may have driven demands for survival strategies, which consequently selected for larger brains, showcasing the impact of global average surface temperatures has on cranial capacity. It then becomes plausible that climate variability may play a role in the evolution of cranial capacitance in humans. However, in contrast to rapid evolutionary change, a study conducted by Sang-Hee Lee and Milford Wolpoff (2003) suggest that evolution of cranial capacity during the Pleistocene period was gradual instead of punctuated. This may suggest that cranial capacity growth follows some sort of linearity growth pattern.

Researchers in the past had concerns over how climate variables may have influenced evolution. During a paleoclimate anthropology conference in November 2005, Richard Potts (2007) wrote what was discussed. The reason for concerns is due to limitations in sampling data, such as Ocean and lake cores provide a more detailed understanding of continuous climate change, while land samples cover smaller areas and are more susceptible to containing discontinuous information because of the locality aspect of land sediments. However land samples are essential in research because they contain fossils, and artifacts of hominins. Findings from researchers have found from comparing climate data is that during times of high variability, Afriancus afarensis went through increased body size changes, while during periods of low variation were points of evolutionary stagnation (Potts 2007). The features used in this paper are time (kya - thousands of years ago), carbon dioxide (parts per million) in the atmosphere, global average surface temperature change (GAST), and the ratio of oxygen-18 to oxygen-16 isotopes (referred to as delta-O-18 or δ18O) to assess past climatic conditions.

One of the objectives of this paper is predicting a continuous variable, making the selected features especially appropriate for linear regression analysis, which excels in modeling relationships involving quantifiable data. Regression analysis is a statistical method used to understand the relationship between a dependent variable and independent variables (Sainani

2013). In this study, the dependent variable is cranial capacity, and the independent variables are carbon dioxide ppm, GAST, and δ18O. Cranial capacity measured in cubic centimeters is a continuous variable which is best predicted using linear regression. Linear regression will predict a value based on an assumed linear relationship between the independent environmental variables. Linear regression will also show how each independent variable affects the dependent variable in a linear fashion. In this way, identifying relationships between each environmental variable is simple, and easy to interpret.

The regression model used in this paper is linear regression using least squares method which minimizes the sum of the squared differences between the observed actual outcomes and the outcomes predicted by the model. Other models that may have been used are decision trees, neural networks, or SVMs (support vector machines), all of which introduce complexity into the model (Golbayani 2020). This hinders the ability to interpret results accurately, and due to the size of the data set may overfit the predictions. Linear regression is a standard approach that optimizes understanding linear relationships between dependent and independent variables (Sainani 2013). Predictions are created using the training data and replaces the actual cranial capacity of records with their corresponding predicted values.

Depending on the range of values seen in each variable and their corresponding coefficients and p-values, we may tell how impactful this variable is towards cranial capacity evolution in humans. In a study conducted by Jessica Ash, she discovered that 52% of variance in cranial capacity could be explained by temperature variations (Ash 2007). An overall evolutionary trend of increasing brain size in primate species as time progresses has been well documented (Neubauer 2012). A reasonable prediction would be a negative coefficient for the time variable (kya) because of how brain size grew to its peak in the present. To calculate the variance explained in each model, looking at the R-squared value for each model explains how much of the variance seen in the dependent variable (cranial capacity for this paper) is explained by the independent variables (environment).

In this study, the predictions presented by linear regression will imply a linear relationship between each environmental variable. The relationship between cranial capacity and each environmental variable will differ based on how each environmental factor changed over the course of human evolution. The coefficients of each variable will tell us the linear relationship of how it affects cranial capacity based on that variable. The p-values will tell the significance of the variable, with a lower value indicating strong significance, and large values as non-significant. Each variable must be evaluated using its p-values to test against the null hypothesis, which represents the probability of incorrectly concluding that there is a dependency when, in fact, no actual effect exists. The p-value needs to be measured below 0.05 in order to consider if something is statistically significant or not. The P-value threshold was chosen based on balancing risks of possibly misinterpreting significance in correlating variables. A P-value threshold of 0.05 is also common amongst research papers and is conventionally standard.

To assess each model, the R-squared values are used to quantify how much variance in the dependent variable—cranial capacity—is explained by the model. R-squared values range from 0 to 1, where a value closer to 1 indicates that a greater proportion of the variance is captured by the model. This metric essentially measures the fit of the model to the observed data. Higher R-squared values signify that the model accounts for a large portion of the variability in the response data around its mean, typically indicating a strong model fit. Conversely, a low R-squared value, approaching 0, suggests that the model fails to adequately explain the observed variability in the data. While there is no universal threshold for an acceptable R-squared value, values above 0.7 are generally considered good, though lower values can still be useful. Because of the confounding nature of cranial capacity evolution, it can be expected that the models in this study will have a low R-squared value.

The linear regression coefficients quantify the impact of each independent variable on cranial capacity. For instance, if the analysis reveals that cranial capacity tends to increase as we approach the present, this would be reflected in a negative coefficient for the time variable. This negative coefficient indicates that as the time variable increases—which represents moving further towards the past in years—cranial capacity actually decreases. This establishes a negative correlation between time and cranial capacity. Essentially, each coefficient in our regression model provides a direct measure of how a one-unit change in an environmental variable influences cranial capacity, allowing us to understand and predict these relationships based on historical data.

An underlying issue presented in this study is the limited carbon dioxide (CO2) data available from Caroyln Snyder's data set (Snyder 2016). In order to address this issue, a subset of fossil records up to 800 kyr of each data set will be tested separately in order to maintain data integrity. A side benefit from doing a subset analysis is identifying how cranial capacitance changes in a short period of time (800kyr). When testing on the entire data sets, the CO2 variable will be omitted from the regression testing and analysis.

MATERIALS

The materials for this study include environment variables: carbon dioxide, global average surface temperature, and oxygen-18 to oxygen-16 isotopes. The materials for cranial capacity are taken from scientific sources whose measurements for hominid skulls could be accurately attained (DeSilva 2021; Ash 2007). The data is then integrated to match corresponding dates between environment variables readings and hominid fossil ages. The data is compiled into separate spreadsheets in order to conduct linear regression learning and view the results. The findings will include p-values, weight coefficients, R-squared values, and graphs detailing the linear relationship between prediction and environment variables.

The first data set is from DeSilva, who collected data from over 900 fossils dating back 7 million years (DeSilva 2021). The oldest fossil mentioned at 7 mya (million years ago) is identified as Sahelanthropus, which is debated amongst anthropologists whether it belongs to the hominin species (Wolpoff 2002). It will be included in this data set, however it could have just as well been dismayed. The rest of DeSilva's data set ranges primarily from 3 mya to present, with a few records over 3 mya. DeSilva compiled their data from published estimates of hominin cranial capacities from scientific literature. The measurements used were cranial volume in cubic centimeters, which is how cranial capacity in humans is measured more commonly.

The second data set was gathered from Jessica Ash in her study on paleoclimate variation and brain evolution in humans. The data set she used is much smaller at 109 fossil records taken from other scientific literature. The measurements used in Ash's data set are averaged by the range of measurements done on each fossil (Ash 2007). This occurs frequently as many scientific literature deploy different measuring techniques which may cause fluctuations in measurements. The oldest fossil in Ash's data set goes up to 2000 kya, and is more widely spread than DeSilva's in terms of dates. DeSilva's data may be larger but a majority of its data comes from very recent fossil records less than 100 years old. In this study, they will be kept separate in order to preserve how each data set trains on linear regression. Then compare findings and interpolate relationships that may be found between them.

The environmental data sets are taken from a study conducted by Carolyn Snyder (Snyder 2016). Measures of GAST are reconstructed using 20,000 sea surface temperature point reconstructions that derive from 59 ocean sediment cores. GAST is then calculated by the difference in temperature found in comparison to modern average temperatures. For example, a -5 value means 5 degrees celsius lower than modern averages. Snyder uses probabilistic simulations from various sources of uncertainty to give an estimate for intervals of which are credible. Linear regression analysis requires independent variables to be continuous values. In order to simplify the model, only a single value for GAST measures is used, which represents the median of the range of a given year interval. In contrast, carbon dioxide and $\delta 18O$ values are taken as is. Each of these environmental features represent a part of the global climate, and together represent an overall global environment. It should be noted however that this is a simplification of the environmental features, and due to using many proxy data may be inaccurate.

DeSilva's and Ash's datasets have some prevalent issues due to GAST recordings only dating back to 2000 kya, and CO2 measurements dating until 800 kya. In order to run linear regression with these missing values, imputation of the missing values was calculated by the mean values of the existing data. Once the model has been trained, the values are removed again and graphed with only the existing values. By doing so it will affect the models performance while removing outlier data that may skew graphs.

Along with having two separate data sets, each is tested on a subset again with fossil records that date up to 800k years ago. This reduces the data set further, however a majority of the data is preserved. Carbon dioxide traps heat in Earth's atmosphere making it a key factor in regulating Earth's climate. It also regulates how plant's photosynthesis rates, which affects food availability (Stockwell 2003). This will change how the data set creates its predictions for cranial capacity, and will need to be compared to the full data sets to better understand how. In total, four different data sets will be used to create separate prediction models using linear regression using linear least squares method.

RESULTS

The data gathered from running linear regression on each of the four data sets show a significant negative correlation between time and the predicted values, with the kyr variable having a p-value of 0.00 indicating statistical significance. Since the time variable is represented inversely starting from the present going further into the past, this negative correlation is better represented as a positive correlation as time progresses towards the present, which is shown in the graphs by inverting the x-axis. This result is not surprising due to cranial capacitance gradually increasing as humans evolved. The other variables listed show the relationship between δ18O, GAST, and co2 vs predicted values gathered from linear regression.

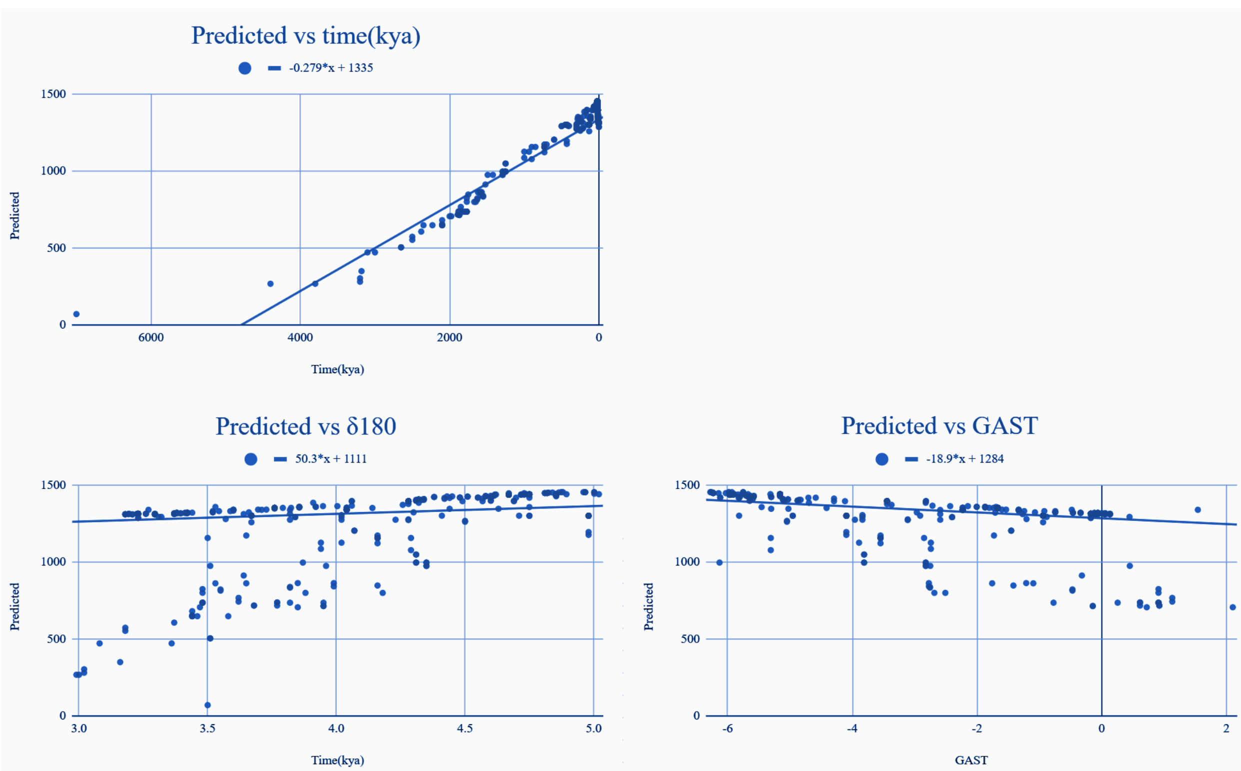

Figure 1.1: Prediction = (-0.327 * kyr) + (49.06 * δ18O) + (-10.55 * GAST) + 1152.69
Tables from DeSilva_2021 full data set. R-squared value of 0.536 which means approximately 53.6% of the variance in the dependent variable "CC" can be explained by the model. Data points for Predicated vs GAST only display up to 2000 kya. Tables represent regression variables against the predicted values from linear least square regression. P-values: kyr -> 0.000, δ18O -> 0.148, GAST -> 0.266.

Using predicted values to understand correlations between variables is a common practice in statistical analysis. Additionally, the weights for each variable is given which tells how the model adjusts its prediction based on the values for each variable. Using both will give a broader picture of the relationship between each environment variable and cranial capacity. The values of the weights coefficients will vary on the impact of which that variable scales. It should be noted that a larger weight value does not mean a stronger correlation as weights are calculated based on the training data and depending on the range fluctuations will greatly affect the values. For instance, δ18O has the greatest weight value in **Figure 1.1**, however the graph shows no correlation between the predicted values and δ18O levels. The weight is large due to how δ18O fluctuates in a range of 3 to 5 which is vastly different from time which fluctuates in a range of 0.1 to 7000 kya.

Figure 1.1 shows little to no correlation between environment variables and predicted values. The weights for the environment variables in the model's prediction show that variables contribute negatively towards the predicted values aside from GAST which holds negative values which impact the predicted values positively. GAST shows the most significant correlation in **Figure 1.1** as GAST levels indicate a downward trend in cranial capacity predictions. This could be understood as GAST decreased through the Pleistocene epoch, human cranial capacity significantly increased. These conditions favored adaptations that enhanced survival and reproductive success across fluctuating habitats, rather than in consistent environments (Ash 2007).

Each model was trained on different data sets, which translates to different learning outcomes from linear regression. The impact of environment variables on cranial capacity showed little to no actual significance except in **Figure 1.2** that resembles DeSilva 800k data set, which showed p-values to all be near zero which means we may reject the null hypothesis, that is, cranial capacity does not depend on environment variables. One caveat to this is the explanation of cranial capacity variability by the environment is relatively low, showing an R-squared value of 0.21, which means 21% of the variability in CC is explained by the model. Comparatively this is the lowest out of the four models in terms of its R-squared value. This shows how complicated cranial capacity is and that the changes are attributed to other factors outside of the environment.

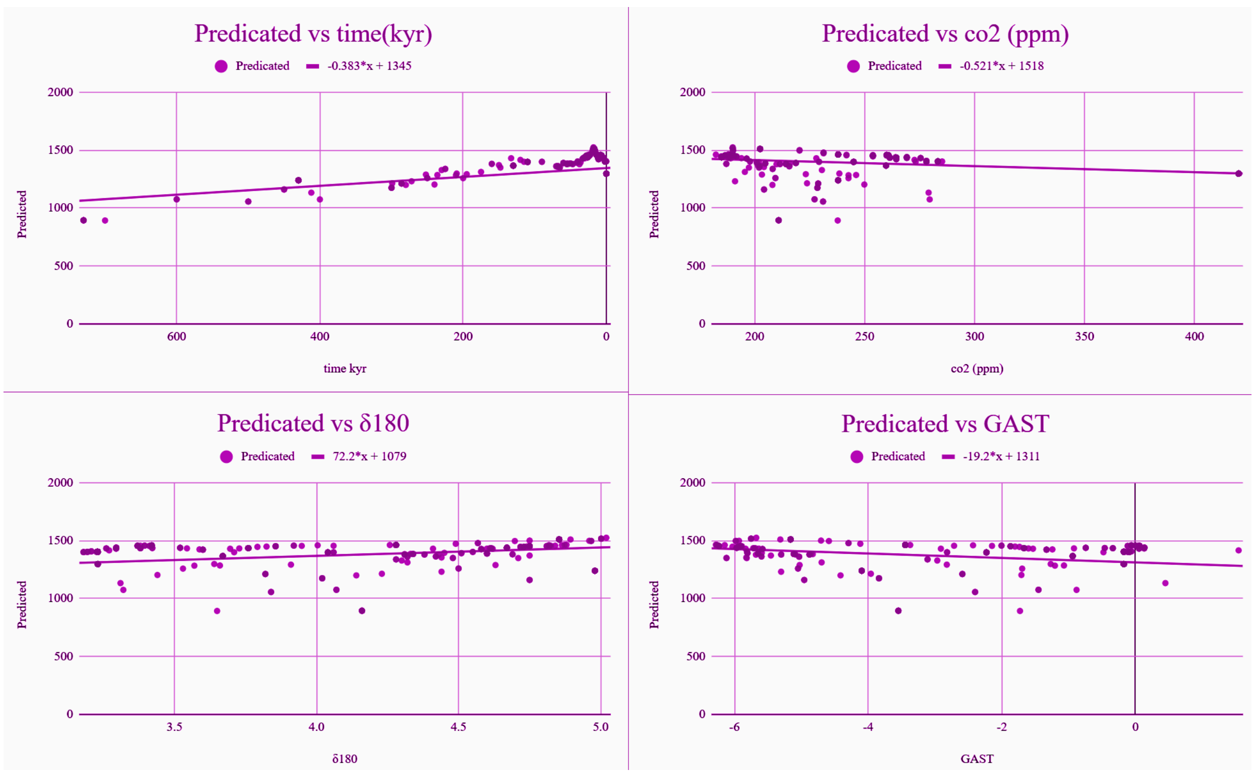

Figure 1.2: Prediction = (-0.78 * Kyr) + (194.91 * δ18O) + (52.14 * GAST) + (-0.78 * CO2) + 1002.7
Tables from DeSilva_2021 800 kya data set. Reduced to better identify relationships between co2 while avoiding imputing missing values. R-squared value of 0.21 which means approximately 21% of the variance in the dependent variable "CC" can be explained by the model. Tables represent regression variables against the predicted values from linear least squares regression. P-values: kyr -> 0.00, δ18O -> 0.00, GAST -> 0.00, CO2 -> 0.00.

Figure 1.2 contains a smaller data set of **Figure 1.1**, consisting of data up until 800 kya. General trends remain similar to the full data set with environment variables showing little correlation with GAST showing the slight downward trend. The purpose of reducing the data set was to hopefully see a better correlation between CO2 and predicted values. However, contrary to this we see little change.

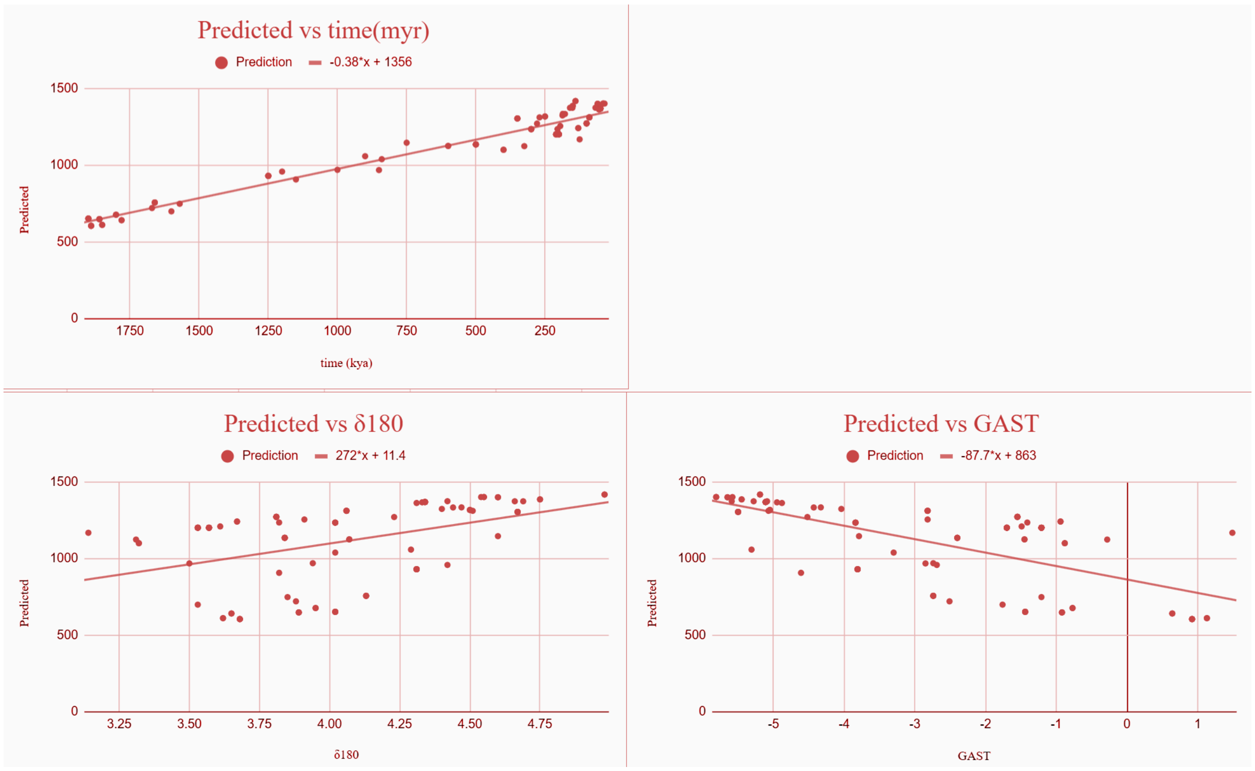

Figure 2.1: Prediction = (-0.36 * Kyr) + (126.27 * δ18O) + (-3.38 * GAST) + 821.58
Tables from Ash_2007 full data set. Omitted co2 in linear regression learning. Tables represent regression variables against the predicted values from linear regression. R-squared value of 0.733 which means approximately 73.3% of the variance in the dependent variable "CC" can be explained by the model. P-values: kyr -> 0.00, δ18O -> 0.131, GAST -> 0.861.

Figures 2.1 and 2.2 correspond towards Jessica Ash's data set of 109 fossil records. The graphs show a significant change in comparison to DeSilva's correlation between environment variables and predicted values. The primary difference between Ash's and DeSilva's data is the distribution of records. Ash's data represents a much wider distribution while DeSilva has a skew in more modern fossils. The wider distribution allows linear regression learning to be less likely to overfit and display a better overall relationship between environmental variables and predicted values.

Figure 2.1 represents the full data set from Jessica Ash's data set. The co2 environment variable is omitted in this data set due to missing values. It was determined that with a small data set and limited available data for co2, it would greatly skew how the data would account for co2 measurements in the case of imputing mean values. The linear regression analysis shows that there are significant positive correlations between δ18O and predicted values. δ18O is an indicator for ice volumes which reflect cooling temperatures (Ash 2007). The impact of δ18O is reflected in the weights as well, with a coefficient of 126.3 (each 1 ppm increase δ18O corresponds to a 126.3 increase in prediction value). The negative correlation in GAST compliments this as lower GAST levels show increases in CC predictions.

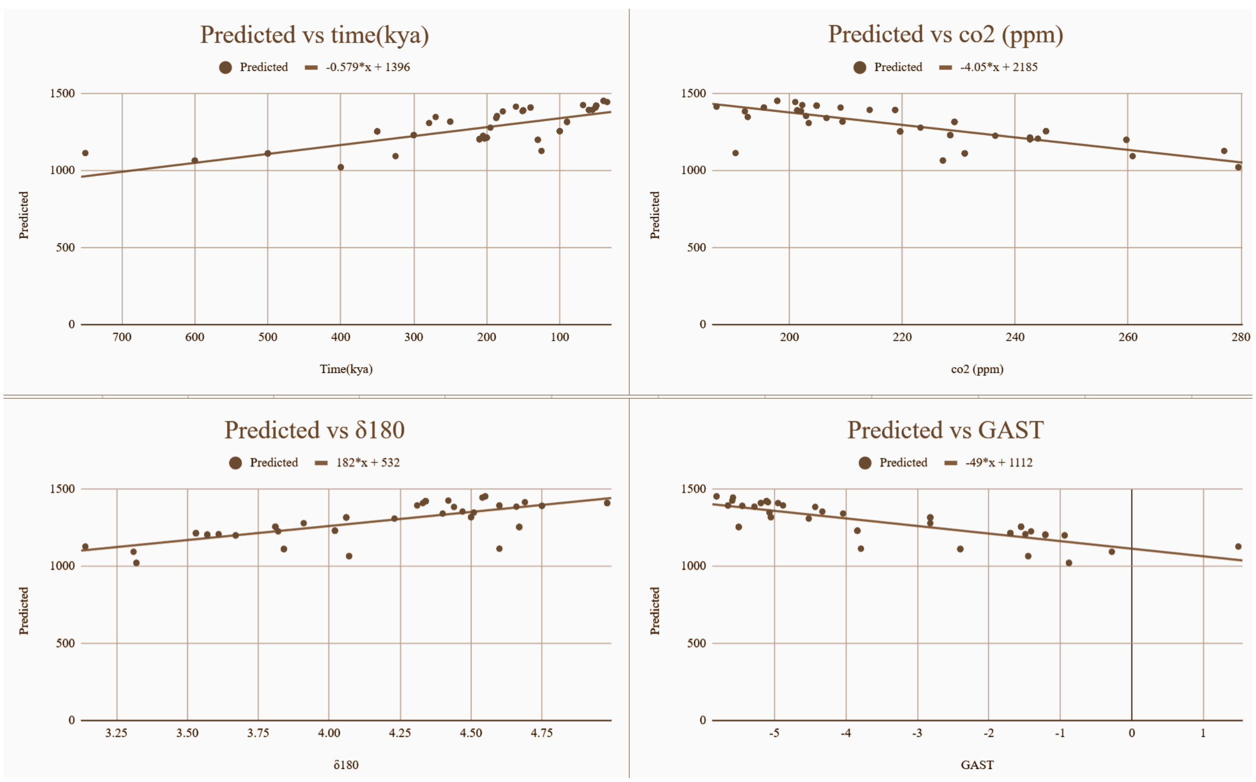

Figure 2.2: Prediction = (-0.47 * Kyr) + (12.15 * δ18O) + (-11.4 * GAST) + (-2.33 * CO2) + 1808.74
Tables from Ash_2007 800 kya data set. Reduced to better identify relationships between co2 while avoiding imputing missing values. Tables represent regression variables against the predicted values from linear regression. R-squared value of 0.321 which means approximately 32.1% of the variance in the dependent variable "CC" can be explained by the model. P-values: kyr -> 0.002, δ18O -> 0.929, GAST -> 0.726, CO2 -> 0.294.

In summary, DeSilva's 800k (**Figure 1.2**) subset data shows the most significant impact of environment variables, having all p-values at 0.00 showing statistically significance. However, the R-squared value of 0.21 indicates that the model does not fit the data well, explaining only about 21% of the variance in cranial capacity. The other data sets (**Figures 1.1, 2.1, 2.2**) did not have P-values for environment variables that indicate statistical significance. These results show that environment variables (used in this analysis) account for 21% of variance in cranial capacity, leaving the rest to confounding factors like diet, social structures, innovations, and many others. It should be noted that linear regression assumes linearity between variables, however research shows that volatile environments accelerate evolution, while periods of stagnant climate reduce evolution change (Davis 2005), which indicates that environmental impact on evolution may not be linear.

DISCUSSION

This paper covers a small area of the possible statistical analysis methods available. One caveat of linear regression is that it assumes a linear relationship between independent and dependent variables. Linear regression is used to represent a predictive model of how a dependent variable (CC) is affected by various independent variables (environment), and also used to predict future values. This is by no means perfect and is generally better used if there actually exists a linear relationship between independent and dependent variables. Researchers generally agree that human evolution is not as simple as a linear relationship, and many factors contribute towards the evolution of cranial capacity.

DeSilva's data set compared to Jessica Ash's has a much different distribution of values, with DeSilva's data containing 985 data points, and Ash's with 109. In DeSilva's data set, over half of records in a small time frame, 0.1 kyr (100 yr), while Ash's data set is more evenly spread ranging from 35 kyr to 1900 kyr. An underlying issue is that the environment variables are derived from the dates in which the fossil records were dated. The environment variables for each of these records are the exact same in value, however the cranial capacity for them are different. The range of cranial capacities seen during this short period ranges from ~907 to ~1786. This can greatly affect how the linear regression model is trained due to the least squares loss, which tries to fit a line that produces least error amongst these values. In contrast, Ash's data set contains a much more evenly distributed distribution of fossil records. The prediction values are much less likely to overfit, and environment values explain a greater portion of the variability found. Ash's full data set has an R-squared value of 0.733 while DeSilva's full data set has an R-squared value of 0.536. Ash's model is able to explain approximately 20% more of the variability seen in cranial capacity.

In **Figure 2.2** there are distinct correlations between each variable and the predicting values. Co2 and GAST show a modest negative correlation, while δ18O show a positive correlation. Based on this model, we may predict a possible future value for experiment. Finding a prediction value is simple, as you just plug in the values into the coefficients for each variable. For example, a time value of -10 (10000 years into the future), δ18O value of 4.5, CO2 value of 440, and GAST of 2 would predict a value of 820.08 CC. The exact same values used in the prediction equation for **Figure 1.1** would give 1107.59 CC. This vast difference comes from the way each prediction model is trained. Because the DeSilva's dataset contains many more data points, it will account for more drastic values reasonably, like CO2, which in the case for **Figure 2.2** shows a dramatic impact on prediction values.

Omitted from this study is the use of a testing set that would give an estimate of the error rate of the linear regression model. The issue with a testing set, is that it would severely limit the already available data as hominin fossil records are scarce. Another problem arises with which data points would be needed to pull from the training set into the testing set. At first, the study included a testing set however quickly found that the error rate was volatile and highly dependent on which data points were chosen. Error rates would fluctuate when using a random 80:20 data split. Seen error rates would fall between 250 to 650 mean square root error. It also would greatly change the prediction model, so for the sake of simplicity and continuity for this study, omitting a testing set seemed reasonable.

With the rise of AI technologies and performance, using neural networks to conduct a similar study to this one would likely perform better. However, the greatest issue when it comes to statistical analysis methods for hominin records is the lack of data. Without an abundance of data points to train a model on, it becomes more likely to overfit and have a large error rate when predicting values. Until more hominin fossil records are uncovered enmasse, neural networks along with other machine learning methods will give subpar results.

The data obtained for this study fails to account for locality of fossil findings. Environment variables are global readings instead of local, which construes the accuracy of exact environment variables that may have played a role in the evolution of cranial capacity in early humans. A more detailed study may contain data for each specific region in which a fossil is found. Doing so would allow for a more detailed account of which environments hominids lived, and better explain the variability in cranial capacity evolution. A great challenge is finding enough fossil records that allows for accurate learning to take place.

REFERENCES

Ash J, Gallup J. 2007. Paleoclimate Variation and Brain Expansion during Human Evolution. Human nature (Hawthorne, NY). 18(2):109–124. doi:10.1007/s12110-007-9015-z.

Beals KL, Smith CL, Dodd SM, Angel JL, Armstrong E, Blumenberg B, Girgis FG, Turkel S, Gibson KR, Henneberg M, et al. 1984. Brain Size, Cranial Morphology, Climate, and Time Machines [and Comments and Reply]. Current Anthropology. 25(3):301–330. doi:10.1086/203138.

Chaline J. 2003. Increased cranial capacity in hominid evolution and preeclampsia. Journal of reproductive immunology. 59(2):137–152. doi:10.1016/S0165-0378(03)00043-3.

Davis MB, Shaw RG, Etterson JR. 2005. Evolutionary responses to changing climate. Ecology (Durham). 86(7):1704–1714. doi:10.1890/03-0788.

DeSilva JM, Traniello JFA, Claxton AG, Fannin LD. 2021. When and Why Did Human Brains Decrease in Size? A New Change-Point Analysis and Insights From Brain Evolution in Ants. Frontiers in ecology and evolution. 9. doi:10.3389/fevo.2021.742639.

Golbayani P, Florescu I, Chatterjee R. 2020. A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees. The North American journal of economics and finance. 54:101251-. doi:10.1016/j.najef.2020.101251.

Hanken J. 1993. Model Systems versus Outgroups Alternative Approaches to the Study of Head Development and Evolution. American zoologist. 33(4):448–456. doi:10.1093/icb/33.4.448.

Lee S-H, Wolpoff MH. 2003. The pattern of evolution in Pleistocene human brain size. Paleobiology. 29(2):186–196. doi:10.1666/0094-8373(2003)029<0186:TPOEIP>2.0.CO;2.

Livingstone DN. 2012. Changing Climate, Human Evolution, and the Revival of Environmental Determinism. Bulletin of the history of medicine. 86(4):564–595. doi:10.1353/bhm.2012.0071.

Mercader J. 2002. Forest people: The role of African rainforests in human evolution and dispersal. Evolutionary Anthropology: Issues, News, and Reviews. 11(3):117–124. doi:10.1002/evan.10022.

Neubauer S, Hublin J-J. 2012. The Evolution of Human Brain Development. Evolutionary biology. 39(4):568–586. doi:10.1007/s11692-011-9156-1.

Potts R. 2007. Paleoclimate and human evolution. Evolutionary anthropology. 16(1):1–3. doi:10.1002/evan.20121.

Rosenberg K, Trevathan W. 2002. Birth, obstetrics and human evolution. BJOG : an international journal of obstetrics and gynaecology. 109(11):1199–1206. doi:10.1046/j.1471-0528.2002.00010.x.

Sainani KL. 2013. Understanding Linear Regression. PM & R. 5(12):1063–1068. doi:10.1016/j.pmrj.2013.10.002.

Snyder CW. 2016. Evolution of global temperature over the past two million years. Nature. 538(7624):226–228. doi:10.1038/nature19798.

Stockwell CA, Hendry AP, Kinnison MT. 2003. Contemporary evolution meets conservation biology. Trends in Ecology & Evolution. 18(2):94–101. doi:10.1016/S0169-5347(02)00044-7.

Sylvester AD. 2006. Locomotor decoupling and the origin of hominin bipedalism. Journal of theoretical biology. 242(3):581–590. doi:10.1016/j.jtbi.2006.04.016.