

Utah State University

DigitalCommons@USU

---

All Graduate Theses and Dissertations

Graduate Studies

---

8-2022

## Implications of Malthus-Boserup Ratcheting for Interpreting the Archaeological Record

Gideon F. Maughan  
*Utah State University*

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>



Part of the [Anthropology Commons](#), and the [Sociology Commons](#)

---

### Recommended Citation

Maughan, Gideon F., "Implications of Malthus-Boserup Ratcheting for Interpreting the Archaeological Record" (2022). *All Graduate Theses and Dissertations*. 8565.

<https://digitalcommons.usu.edu/etd/8565>

This Thesis is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).



IMPLICATIONS OF MALTHUS-BOSERUP RATCHETING FOR INTERPRETING

THE ARCHAEOLOGICAL RECORD

by

Gideon F. Maughan

A thesis submitted in partial fulfillment

of the degree

of

MASTER OF SCIENCE

in

Archaeology and Cultural Resource Management

Approved:

---

Jacob Freeman, Ph.D.  
Major Professor

---

David Byers, Ph.D.  
Committee Member

---

Judson Finley, Ph.D.  
Committee Member

---

D. Richard Cutler, Ph.D.  
Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY  
Logan, Utah

2022

Copyright © Gideon Maughan 2022  
All Rights Reserved

## ABSTRACT

Implications of Malthus-Boserup Ratcheting for Interpreting the Archaeological Record

by

Gideon F. Maughan, Master of Science

Utah State University, 2022

Major Professor: Dr. Jacob Freeman  
Department: Sociology, Social Work, and Anthropology

An exponential-like pattern of population growth has been documented across North America, but variations of this pattern at the regional level have yet to be widely explored. The field of archaeology presently lacks a theory that explains the diversity of long-term population-growth trajectories that human societies might exhibit. I propose the use of an established Malthus-Boserup population ratcheting model to explore the variations in population-growth trajectories. The Malthus-Boserup ratcheting model describes population wide responses to demographic pressure through changes in behavior or strategy that increase regional demographic saturation. To model Malthus-Boserup ratcheting, I generate simulated bi-logistic curves. I then run exponential regressions on the curves to develop expectations for comparing exponential regressions with bilogistic regressions taken on Summed Probability Distributions from radiocarbon datasets. Using these expectations, I hypothesize that the size of changes in demographic saturation result in regional differences in exponential growth rate. I use radiocarbon datasets from four regions in North America. I perform a cluster analysis on each summed probability distribution to determine how many waves of growth to include in

each multilogistic regression. Using R, I generate both exponential and multilogistic regressions for each dataset. I then use the Bayesian Information Criterion to compare data loss associated with each model and discuss the implications of each model's parameters and their meanings in the archaeological record. Using the Malthus-Boserup model to describe the radiocarbon record is more informationally conservative than the exponential model, both in terms of data description and interpretation.

(101 pages)

## PUBLIC ABSTRACT

Implications of Malthus-Boserup Ratcheting for Interpreting  
the Archaeological Record

Gideon F. Maughan

Prehistoric populations across North America seem to grow exponentially, with some variation between regions. Archaeologists have explored the differences somewhat, but have not explained the differences or the sustained growth with any reference to what may be going on under the surface in a way that is relevant to all regions. I propose that environmental limits on population are shaped by what populations eat and how they acquire food, and that when populations are large enough to feel the scarcity in their environment, they change their way of life in a way that increases those limits. The model I propose is well-established, and is called the Malthus-Boserup ratcheting model. I mathematically describe the Malthus-Boserup ratcheting model using a fixed population growth rate and an array of changes in environmental limit, both in the amount of change and rate of change (or how quickly my imaginary populations change their way of life). I then simplify my descriptions of these curves using exponential curves, as archaeologists might simplify real population growth curves in much the same way. I compare the models with their exponential descriptions to form expectations for what the values in exponential curves might mean regarding the archaeological record. Despite having set a constant population growth rate, the exponential curves grow at different rates, depending primarily on the amount of change in environmental population limits. I hypothesize that a list of regions assembled in order of largest to smallest change in population limits

should match a list of the same regions assembled in order of exponential growth rate. I use tools in R statistical software to describe the population growth curves in four regions using both exponential curves and summed logistic curves. I then arrange the list of regions according to both the growth rate of the exponential curves and the change in the limits found in the logistic curves. The lists do not match, which suggests that, for a variety of reasons, the exponential curves do not adequately describe the underlying Malthus-Boserup ratcheting process. I then compare the models using the Bayesian Information Criterion. Bayesian Information Criterion is an indicator that increases with both the unexplained variation in the dependent variable and the number of explanatory variables used. In this sense, when comparing models describing the same data, lower values suggest either that information retained outweighs a model's complexity, or a model's simplicity outweighs the amount of information lost. In all four cases, the information retained in the Malthus-Boserup ratcheting model outweighs the model's complexity. Furthermore, the summed logistic models have parameters that researchers can interpret beyond simple rates of change.

## ACKNOWLEDGMENTS

There are so many people without whom I could not have done this work. Erick Robinson, thank you for entrusting me with the data. McKinley Benson and Megan Sills, you have both been incredible in helping me stay on top of dates and forms. Megan, thank you especially for your help with boiling this thesis down to its more important points. Dave, thank you for being there during my most frustrating times and letting me cry in your office. Dr. Freeman, words cannot express how indispensable you have been to me as an advisor. Your patience and assistance in keeping my thesis on topic, as well as your help in understanding the math have been invaluable. Judson, thank you for your time serving on my committee. The enthusiasm and interest you expressed in responding to my request helped me remain positive in a difficult time. Darcy, Cayla, and Paige, thank you for discussing problems I encountered and for just being there. To my wife Tabitha, thank you. You have carried the lion's share of our burdens from the usual housework to our relocation. I could not have done this without your patience and kindness.

Sincerely,  
Gideon F. Maughan



## CONTENTS

Page

ABSTRACT.....	iii
PUBLIC ABSTRACT .....	v
ACKNOWLEDGMENTS .....	vii
LIST OF TABLES.....	x
LIST OF FIGURES .....	xi
CHAPTER 1 INTRODUCTION.....	1
CHAPTER 2 THEORY AND HYPOTHESES.....	3
CHAPTER 3 DATA AND METHODS .....	19
CHAPTER 4 EMPIRICAL RESULTS .....	34
CHAPTER 5 DISCUSSION AND CONCLUSION .....	39
APPENDIX I: SPDS WITH EXPONENTIAL REGRESSIONS.....	50
APPENDIX II: CLUSTER ANALYSES .....	54
APPENDIX III: MULTILOGISTIC CURVES WITH CONFIDENCE ENVELOPES ..	62

APPENDIX IV: R CODE.....66

## LIST OF TABLES

	Page
Table 3.1 Hopkins Indices report degree of clustering, ranging from 0.5 to $< 1$ .....	27
Table 3.2 Best fit estimation used for each wave .....	32
Table 4.1 Exponential growth rates with confidence envelope .....	34
Table 4.2 Rank order of regions based on exponential growth .....	35
Table 4.3 Multilogistic regression results.....	35
Table 4.4 Relative change in saturation level from multilogistic regression.....	37
Table 4.5 Regions rank ordered by relative growth in saturation.....	38
Table 4.6 BIC values for exponential and multilogistic regressions .....	38

## LIST OF FIGURES

	Page
Figure 2.1 Malthus, Boserup, and Malthus-Boserup Ratcheting.....	4
Figure 2.2 Exponential curves fit to bilogistic growth models, four by four.....	11
Figure 2.3 A heatmap of exponential curves .....	13
Figure 2.4 Change in growth rate (y axis) by change in $\alpha$ (x axis) at $kg = 1$ . .....	14
Figure 2.5 Exponential curves fit to bilogistic growth models, 80 time steps.....	15
Figure 2.6 Change in growth rate by change in $\alpha$ , 320 time steps .....	16
Figure 2.7 Exponential curves fit to bilogistic growth models, 320 time steps.....	16
Figure 3.1 Sampling units selected for this study .....	22
Figure 3.2 SPD for Nevada with confidence envelope for exponential regression .....	26
Figure 3.3 Cluster Plot with ellipsoids for Texas .....	29
Figure 3.4 SPD for Texas, colored by cluster .....	29
Figure 3.5 Cluster Plot with ellipsoids for PNW .....	30
Figure 3.6 SPD for PNW, colored by cluster.....	30
Figure AI.1 Exponential regressions for Texas SPD and confidence envelope .....	50
Figure AI.2 Exponential regressions for Nevada SPD and confidence envelope.....	51
Figure AI.3 Exponential regressions for Wyoming SPD and confidence envelope.....	52
Figure AI.4 Exponential regressions for PNW SPD and confidence envelope.....	53
Figure AII.1 Cluster analysis for Texas.....	54
Figure AII.2 SPD for Texas, colored by cluster .....	55
Figure AII.3 Cluster analysis for Nevada .....	56
Figure AII.4 SPD for Nevada, colored by cluster.....	57

Figure AII.5 Cluster analysis for Wyoming ..... 58

Figure AII.6 SPD for Wyoming, colored by cluster ..... 59

Figure AII.7 Cluster analysis for PNW..... 60

Figure AII.8 SPD for PNW, colored by cluster ..... 61

Figure AIII.1 Bilogistic curve with confidence envelope fit to Texas ..... 62

Figure AIII.2 Bilogistic curve with confidence envelope fit to Nevada..... 63

Figure AIII.3 Bilogistic curve with confidence envelope fit to Wyoming..... 64

Figure AIII.4 Multilogistic curve with confidence envelope fit to PNW ..... 65

## CHAPTER 1 INTRODUCTION

Building sustainable societies requires protecting the earth's biophysical systems from degradation (Biermann et al., 2012; Catton, 1987; Cohen, 1995; Fedoroff and Cohen, 1999; Hardin, 1968; Rockström et al., 2017). The most critical process that impacts such systems is the long-term growth of human populations (Catton, 1987; Dietz et al., 2003). Yet, we still lack a theory that explains the diversity of long-term population growth trajectories that human societies exhibit. In part, this is a data issue: Most studies focus on global population data over the last 500 years or a few select regions with good data (Cohen, 1995; Kelly et al., 2013; Lucarini et al., 2020; Meyer and Ausubel, 1999; Palmisano et al., 2018; Pardo-Gordó et al., 2019; Riede, 2009). However, through the recent development of large data sets useful for estimating paleo population attributes (e.g., Martindale et al., 2016), archaeologists now have the opportunity to both catalog the diversity of growth trajectories that human societies have followed over the last 12,000 years and develop explanations for this underlying diversity (Bird et al., 2020; Chaput et al., 2015; Chaput and Gajewski, 2016; Freeman et al., 2018a, 2018b; Johnson and Brook, 2011; Peros et al., 2010; Robinson et al., 2019; Shennan et al., 2013; Zahid et al., 2016).

Specifically, in this thesis, I ask: *Do hunter-gatherers experience the same long-term trajectories of population growth and growth rates regardless of ecological context?* To help answer this question, I present an iterative model of logistic population growth in which the carrying capacity of an area changes over time as a function of resource substitution (Freeman et al., 2021; Meyer et al., 1999). I then evaluate the predictions of the model using radiocarbon records to compare the growth trajectories of

hunter-gatherers in western N. America, and I end with a discussion of the potential and limitations of radiocarbon records for understanding the diversity (or homogeneity) of human population growth throughout the Holocene.

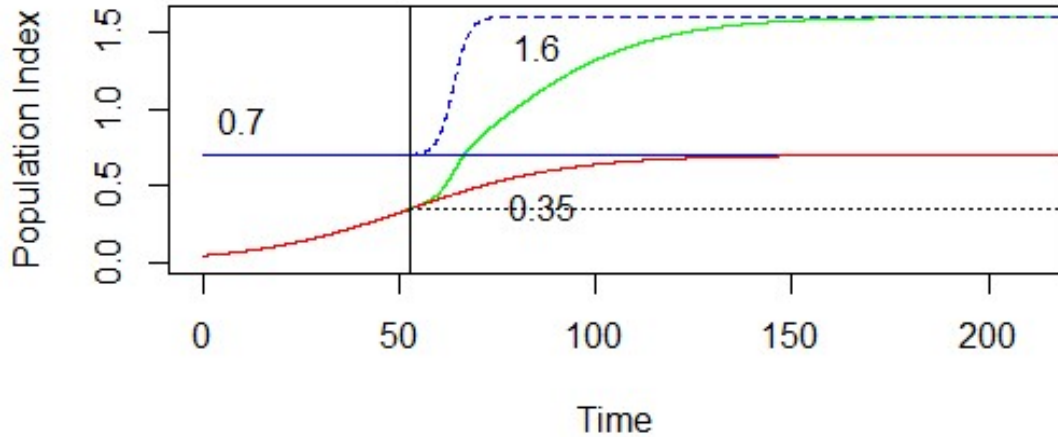
The basic conclusion of my thesis is that while the radiocarbon record may take the appearance of exponential growth, multilogistic curves may be more appropriate than exponential curves in describing and exploring demographic change. The exponential model assumes unchecked growth, ignoring population limitations imposed by human-environment interactions. While the exponential model is more parsimonious, interpreting its parameters is complicated. The coefficient for left or right translation is presented as a mathematical convenience, and has no explicit connection to demographic theory. The Malthus-Boserup ratcheting model provides a theory-based approach to understanding demographic curves. Mathematical models produced with Malthus-Boserup ratcheting in mind use parameters that relate to demographically relevant theoretical entities such as population growth, saturation level, time of inflection, nominal growth rate, and initial population. Finally, the degree to which summed logistic curves conserve data in its description outweighs model complexity to a greater extent than exponential curves' parsimony outweighs their data loss. Fitting multilogistic models to archaeological records provides a method for estimating levels of population pressure in archaeological contexts, and such models provide hypotheses for explaining patterns of social and technological change in relationship to changes in macro-demography.

## CHAPTER 2 THEORY AND HYPOTHESES

The inspiration for my thesis is a well-documented pattern, where almost all radiocarbon records from western N. America display long-term exponential increases over the last 12,000 years (Bird et al., 2020; Freeman et al., 2018a, 2021). It is common, thus, to fit null exponential models to these records. However, it is not clear why fitting an exponential model is appropriate or how one might interpret differences in such models across cases. In my thesis, I propose that the apparent long-term exponential increase in radiocarbon records may result from multiple regimes of density-dependent population growth. My main goal here is to describe a bi-logistic model that formalizes this proposal and then plot the results of the ratchet model of population growth. By conducting simulations, I both explore the complexity of interpreting the fit an exponential model fit to data generated by the bi-logistic process and how the long-term growth trajectories and rates of hunter-gatherer populations might vary based on ecological context. The ratchet model might provide a descriptive explanation for why areas may display shifts between different regimes of population growth.

This model was first developed by Meyer and Ausubel (1999) and adapted to hunter-gatherers by Freeman et al. (Freeman et al., 2021). The model combines three assumptions from Malthusian theory on population growth and Boserupian theory on agricultural intensification. Figure 3.1 summarizes the theories of Malthus and Boserup





*Figure 2.1 Malthus, Boserup, and Malthus-Boserup Ratcheting* Malthus argued that populations grow geometrically (red curve) under ideal circumstances until each additional person reduces resources available to the average person (black vertical line and black dotted line). After this point, Malthus argued, populations enter a zone of pressure where the lack of available resources impacts the population's quality of life until the mortality rate matches the birth rate. The result is a population resting at saturation (blue solid line). Boserup argued that as populations grow, quality of life decreases. When quality of life falls below a culturally accepted threshold, people adopt resource intensification strategies and quality of life improves. The Malthus-Boserup ratcheting model combines Malthusian population dynamics with Boserupian intensification. Under ideal conditions, populations experience exponential-like growth. Labor inputs have decreasing marginal returns. Quality of life decreases in the zone of population pressure as the population approaches saturation. When quality of life decreases below some culturally acceptable threshold, the society adopts a new resource intensification strategy. This intensification results in a change in saturation level (blue dashed curve) and a new population growth curve (green curve). In this way, the population ratchets itself up with an increasing saturation level.

(1) Populations unconstrained by competition grow exponentially:

$$p(t) = \beta e^{rt} \quad (2.1)$$

where  $\beta$  is the initial population size;  $r$  is the maximum growth rate of the population;

and  $t$  is time.

(2) Increases in food production, at best, grow linearly over the long-term, and, thus, competition for resources eventually limits population growth. These first two assumptions are Malthusian principles (Malthus, 2003), and the logistic model is the classic function that captures both principles.

(3) A zone of population pressure exists in any Malthusian system between a culturally defined level of well-being and the Malthusian equilibrium where fertility and mortality are balanced (Wood 1998, 2020). Well-being is best thought of here as an expected flow of resources determined by cultural and ecological constraints. In this zone of population pressure, the fragility of the demographic system sends signals (individuals have trouble meeting their expected flow of resources over some time interval). In turn, the signals create an environment that favors the substitution of one modal social and technological strategy for another with a higher carrying capacity (Freeman et al., 2021; Wood, 1998, 2020). This may occur through a process of innovation or may just entail the reorganization of existing knowledge sets, rules, and time allocated to resources already used, or this process may not occur at all, in which case a population system would enter a 'Malthusian Trap.'

For example, on a small scale, a person visiting a resource patch may find that, as a result of demographic pressure, this resource patch is not as fruitful as usual and either gathers fewer resources for the day, moves to another patch, gathers for longer, or attempts to increase the productivity of the patch through landscape modification (e.g., diverting a stream). Any of these strategies may represent a deviation from this person's usual gathering activities. The change in strategy increases the saturation level of the

patch, scaled by the number of people adopting the new strategy. The number of people adopting the new strategy can grow quickly—for example, everyone visiting a resource patch finds that they must gather for longer or visit a different resource patch—or slowly. The reasons why people might adopt a strategy more slowly are somewhat varied. For example, new strategies may be relatively inaccessible—the availability of suitable resources for inclusion or substitution in a diet may be lower in sparse environments—or undesirable—the new strategy includes consuming less desirable food, such as foods the society deems “gross,” “dirty,” or even “sacred,” or foods that require a greater input in time and effort, such as foods that require digging an irrigation channel. Whatever the rate, the proportion of individuals participating in the new strategy grows, and with it, the population saturation level grows until the saturation level equilibrates at the maximum level achievable given the new strategy.

To capture the assumptions above with a descriptive growth model, I rely on a simple modification to the logistic function used by Meyer and Ausubel (1999) and Freeman et al. (2021). Formally,

$$\dot{p} = pr \left( 1 - \frac{p}{k(t)} \right) \quad (2.2)$$

where  $p$  is population size at time  $t$ ;  $r$  is the maximum growth rate; and  $k(t)$  is the variable carrying capacity of an environment—what we will call population saturation from here on—that depends on the modal social and technological strategies of a population at time  $t$ . Here  $r$  is a constant, regardless of technological differences, reflecting the results of (Zahid et al., 2016) that human populations may have a constant,

biologically limited rate of growth. The key idea in equation (2.2) is that population saturation changes as a function of the modal subsistence strategy at time  $t$ . This means that changes in social and technological organization affect growth processes by changing  $k$ . As noted above, this change in  $k$  might occur due to innovations, adopting new strategies via diffusion, or through a simple process of substitution. The most likely process, in our view, among hunter-gatherers is resource substitution.

For example, ethnographic data illustrate the positive effect of net primary productivity on human population densities (Binford, 2002; Freeman et al., 2020; Tallavaara et al., 2018). However, the effect is quite small compared to differences between societies primarily dependent upon agriculture and those dependent upon hunting and gathering. The mean population density of hunter-gatherer societies is about 300 percent lower than the mean population density of agricultural societies at the same level of NPP (Freeman et al., 2020: Figure 4). It takes an increase in NPP of about 1000 grams per square meter per year for the mean hunter-gatherer population density to increase 300 percent (an equivalent in geographic space of moving from Bozeman, Montana to Huston, Texas in North America).

I do not suspect that the reader will find anything controversial about the relationship between more reliance on resources with a higher energy density per unit area and higher population saturation values among human societies. Any objection, I suspect, would come from the assumption that the substitution of one set of resources for another occurs in response to instabilities in a production system caused by entering a zone of population pressure. Understanding why this may or may not occur requires

building dynamic models that include individual decision making and system level components (see e.g. Freeman et al., 2019; Puleston et al., 2014 for beginning frameworks). The strength of the descriptive model here is that we can evaluate whether the population growth trajectories of prehistoric regions are consistent with a population ratchet assumption in which  $k$  increases sequentially through time following periods of population pressure or not.

Descriptively, we can write the change in  $k$  as a logistic function:

$$\dot{k} = \alpha(k - k_{si}) * [1 - (k - k_{si})/k_g] \quad (2.3)$$

Where  $\alpha$  is the maximum rate at which agents substitute one modal strategy for another, and  $k_g$  is the interval between  $k_{s1}$  and  $k_{s2}$ . For example, if  $k_{s1} = 0.75$  and  $k_g = 1$ , then  $k_{s1} + k_g = k_{s2} = 1.75$ .

We define  $t_{pi}$  as the time at which a growing population reaches demographic pressure, thus we can write the solution to equation (2.3) as

$$k(t) = k_{si} + \frac{k_g}{1 + e^{-\alpha(t-t_{pi})}} \quad (2.4)$$

where  $t$  is time.

Finally, we define  $p_0$  as the initial population density of a given area and write the solution to equation (2.2) as

$$p(t) = \frac{k(t)}{1 + \left(\frac{1}{p_0/k(t)}\right)e^{-rt}} \quad (2.5)$$

We can generalize equation  $p(t) = \frac{k(t)}{1 + \left(\frac{1}{p_0/k(t)}\right)e^{-rt}}$

(2.5) for more than two  $k$  if needed; however, two  $k$  are sufficient to illustrate the general dynamics. The power of this simple, bilogistic model is that I can systematically analyze how changes in the adoption of a new technology and the relative change in population saturation impact trajectories of long-term population growth in terms of the shape of a growth curve and the mean rate of change experienced in a given area.

Note here that the above model attempts to describe population growth as process driven by changes in competition, technology, and social organization. I assume that external processes, like climate change, that might impact  $k$  in the real world are effectively constant in this model. This is not to say that climate change is unimportant; just that to develop a theoretical understanding of the different shapes and long-term growth rates that human populations might exhibit, it is useful to simplify and first understand the dynamics of this model free of climate perturbations to  $k$ . Future research can easily add climate parameters to perturb  $k$ , positively or negatively, if this first approach helps clarify the problem.

### **Model Insights and Hypotheses**

In this section, I run simulations to explore how changing the parameters of the above model affects the shape of the long-term population growth curve of human

societies. Using equation  $p(t) = \frac{k(t)}{1 + \left(\frac{1}{p_0/k(t)}\right)e^{-rt}}$

(2.5), I calculate a demographic curve for a fixed saturation level ( $k = 0.75$ ,  $kg = 0$ ,  $r_1 = 0.2$ ) and find the earliest instance where the population is at least 50% saturation.

This is the time at which the modeled population begins experiencing the kind of feedback from the environment that signals demographic pressure as outlined above. I use this value ( $t = 17.01$ ) as an estimate for the  $t_{pi}$  parameter in equation  $k(t) = k_{si} +$

$$\frac{k_g}{1+e^{-\alpha(t-t_{pi})}}$$

(2.4)—representing in the model the earliest behavioral response to environmental feedback. To numerically and visually illustrate the bilogistic model's dynamics, I generate datasets for four values of  $\alpha$  and four values of  $k_g$  (sixteen data sets) using

$$\text{equation } k(t) = k_{si} + \frac{k_g}{1+e^{-\alpha(t-t_{pi})}}$$

(2.4), representing four rates of adoption and four maximum amounts of change in population saturation level respectively, as well as the demographic response to these changes.

I run exponential regressions on each of these sixteen data sets, and plot the demographic curve and exponential regression in a plot matrix in order to visualize how differences in the amount of change in saturation level and differences in the rate of adoption of a new strategy impact both the shape of the demographic curve and growth rate of an exponential regression fit to a bilogistic curve. As noted above, I fit the exponential model to understand of how an underlying population process that displays multiple  $K$  might impact the interpretation of a logistic model fit to such simulated data.

I used the exponential function

$$f(t) = e^{(a+bt)}$$

(2.6)

where  $t$  is time. The point of this exercise is to assess how the slope of an exponential regression, a simple descriptive model often used by archaeologists to describe real world radiocarbon trends, might vary as a function of the rate of adopting new technologies and the change in population saturation of an environment as described by the bilogistic model.

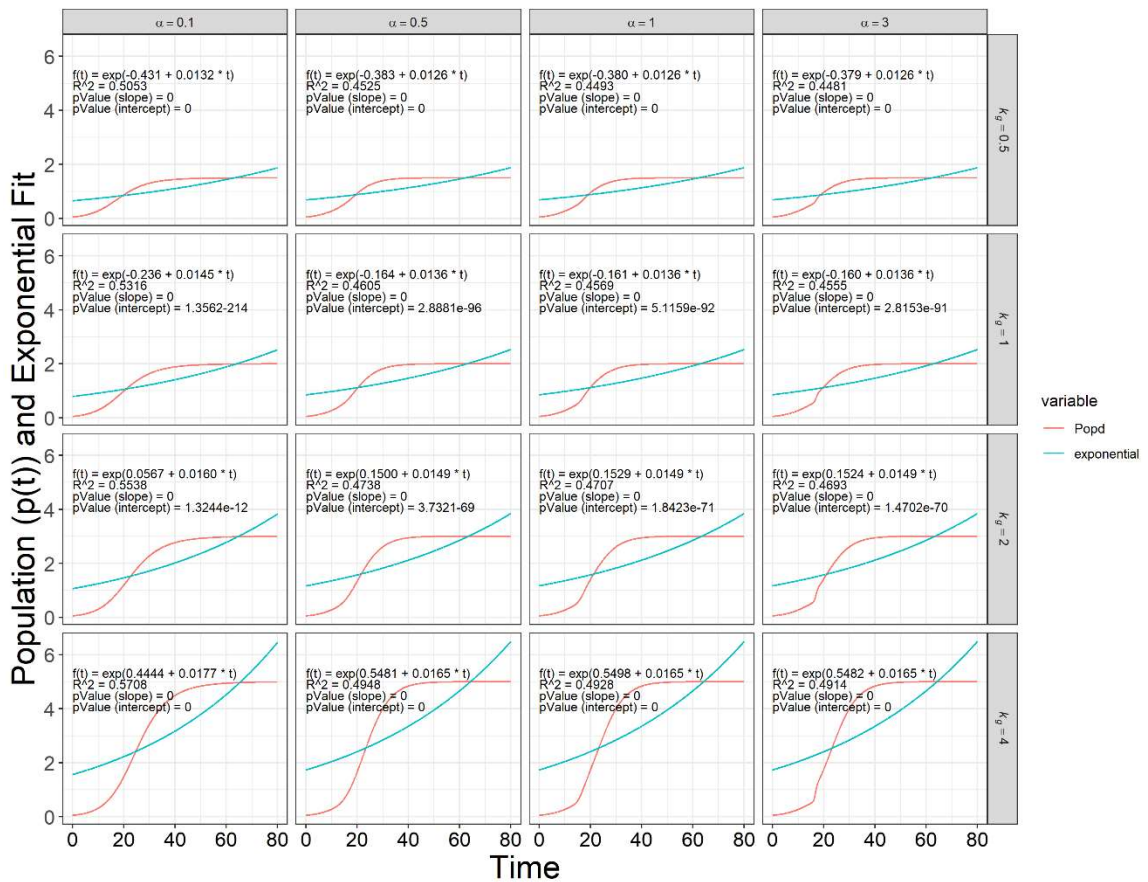


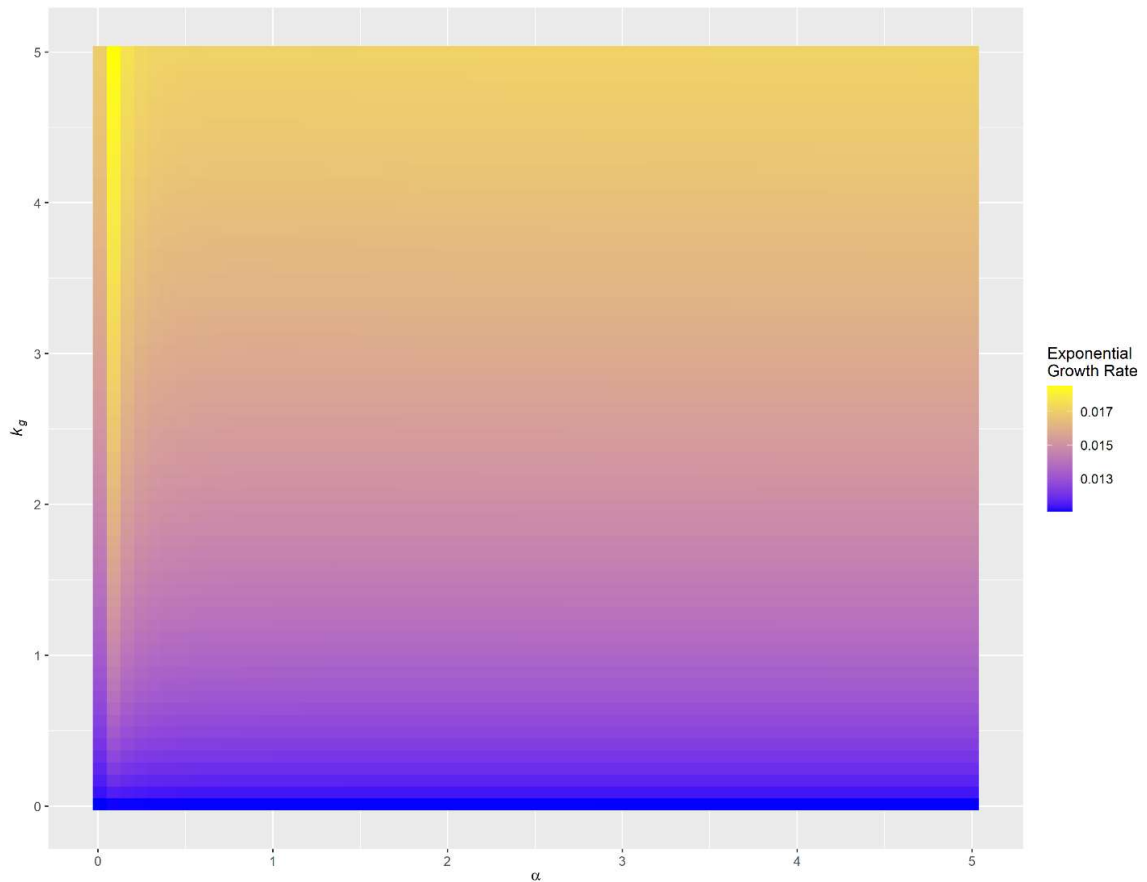
Figure 2.2 Exponential curve fit to bilogistic growth model for four values  $\alpha$  and four values  $k_g$ . The plot illustrates how differences in values  $\alpha$  and  $k_g$  change the shape of the simulated bilogistic curves (red curves). The blue curves illustrate exponential models statistically fit to the underlying bilogistic growth curves. This allows us to



*observe how estimated growth rates from exponential models might vary in terms of estimated fit and parameters when fit to data that result from multiple waves of density dependent population growth.*

By looking at the plot matrix (Figure 2.2), I see that the differences in  $k_g$  have a more profound effect on the exponential growth rate than do differences in the rate at which the new strategy is adopted. That is, if a new modal form of social and technological organization increases population saturation by 400%, then the estimated exponential growth rate of that curve is faster than when  $k$  increases by 300, 200, or 100%. Differences in the rate at which a new strategy is adopted do appear, however, to have an influence on the apparent slope of the demographic curve, particularly from the time at which the new behavior is adopted to the time at which the curve reconciles with the second phase of growth. This, in turn, affects both the horizontal translation—left or right shift, which would describe the portion of the exponential curve that best fits the population curve—and exponential growth rate of the fitted curve. This raises the question of whether values of  $\alpha$  within reasonable ranges of  $\alpha$  and  $k_g$  can produce sufficient results as to render uncertain claims as to whether ranked differences in exponential growth rate arise from differences in the rate of adoption of a new strategy or differences in the amount of change in saturation level.

To assess the impact of rate of adoption, I plotted a heat map of the growth rates of exponential regressions on modeled bilogistic curves for 64 values of  $\alpha$  ranging from 0.01 to 5 and 64 values of  $k_g$  ranging from 0.1 to 5.

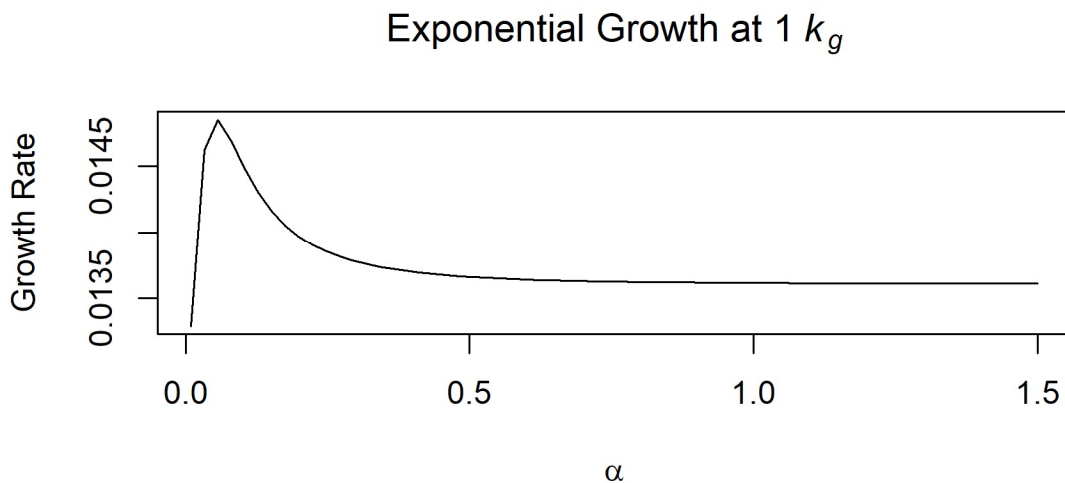


*Figure 2.3 A heatmap of exponential curves This heatmap shows how  $\alpha$  (increasing from left to right on the x axis) and  $k_g$  (increasing from bottom to top along the y axis) affect the growth rate of exponential curves regressed on population growth given a logistic upward shift in saturation level. Each pixel is colored according to the value of the exponential growth rate observed at each pair of  $\alpha$  and  $k_g$ . Higher exponential growth rates appear in yellow, while lower rates appear in blue, and median rates appear in purple. Note the general trend from blue through purple to yellow with increasing  $k_g$ , with a somewhat more rapid change for a band of lower  $\alpha$  values.*

From the heatmap, it appears that (1) as expected, increases in  $k_g$  increase the growth rate of the exponential curve fit to the bilogistic model. (2) The growth rate

increases sharply as  $\alpha$  increases from 0.01 to 0.05, and decreases with increasing  $\alpha$  thereafter. This is particularly relevant to higher values of  $k_g$ , but still notable at low  $k_g$  values.

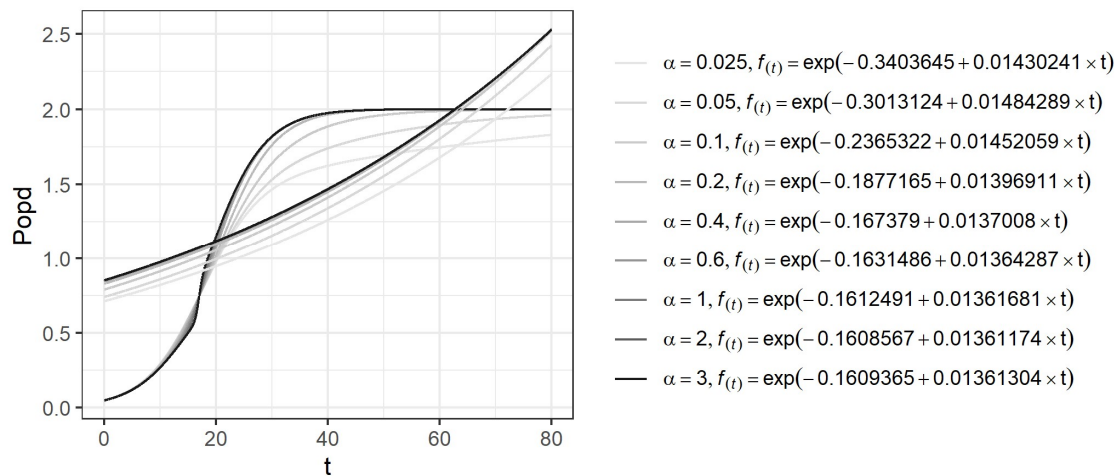
To visualize the effect this narrow band of  $\alpha$  values may potentially have on growth rates reported by exponential regression and to better understand the nature of this effect, I plotted a cross-section of growth rates for 64  $\alpha$  values between 0 and 1.5 where  $k_g = 1$ .



*Figure 2.4 Change in growth rate (y axis) by change in  $\alpha$  (x axis) at  $k_g = 1$ . This graph represents a cross-section of the heatmap in Figure 2.3, truncated at  $\alpha = 1.5$ . With increasing  $\alpha$ , the exponential growth rate changes with logarithmic-like growth, followed by an exponential decrease.*

Figure 2.4 offers a finer-grained understanding of the range of  $\alpha$  values on estimated growth rates in the bilogistic model. In Figure 2.5, I plotted nine more curves with exponential regressions for one value of  $k_g$ : three from near the spike, three from

the sharply decreasing portion of the curve, and three from the mostly flat portion of the curve. For a better understanding of the range and nature of the effects of  $\alpha$  on the translation coefficients and growth rates of exponential regressions, I modeled an exponential curve fit to the demographic curve resulting from several  $\alpha$  values as above, but for a single  $k_g$  value, plotting the results on a single chart.



*Figure 2.5 Exponential curves fit to bilogistic growth model for nine values  $\alpha$  and one value  $k_g$ , 80 time steps*

Right away, I could see that the curves generated in the lower ranges of alpha fail to reach equilibrium within the 80 time-steps. The result is somewhat complicated. First, the late end point of the curve I am feeding to the exponential regression is lower, and the regression is blind to whether the curve has truncated early or reached  $k$ .

Counterintuitively, the rate of exponential growth increases rather than decreases as the later end of the exponential curve shifts downward with the demographic curve. This is because the points over the interior portion of the exponential curve are lower and lack the upper elbow, so the exponential curve is deeper. To capture the decrease in rise

over run slope with the increased belly depth, the exponential regression requires a higher growth rate with greater horizontal shift to the right. To know how this effect is driven by whether and how long the demographic curve equilibrates, I generated the two previous plots, this time with 320 time-steps rather than 80 (Figures 2.6 and 2.7).

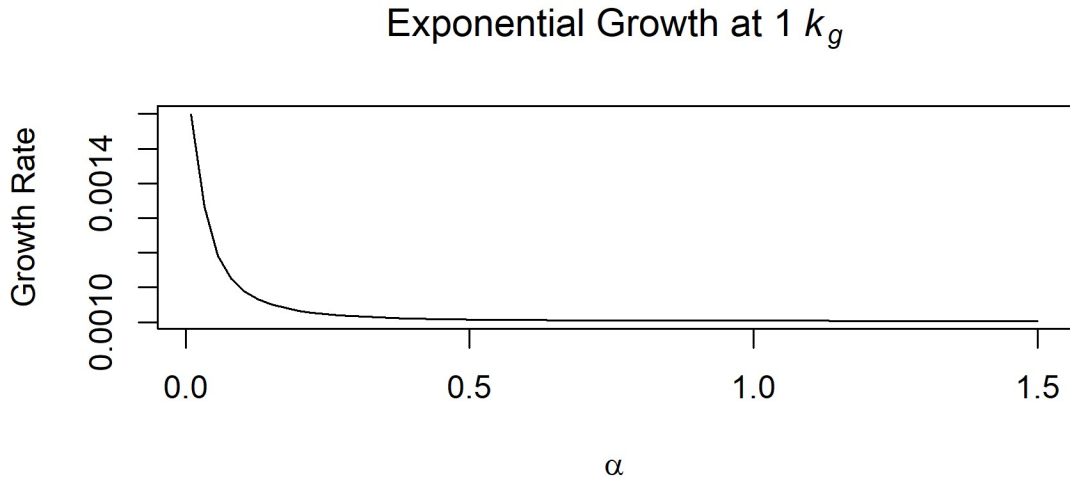


Figure 2.6 Change in growth rate by change in  $\alpha$ , 320 time steps

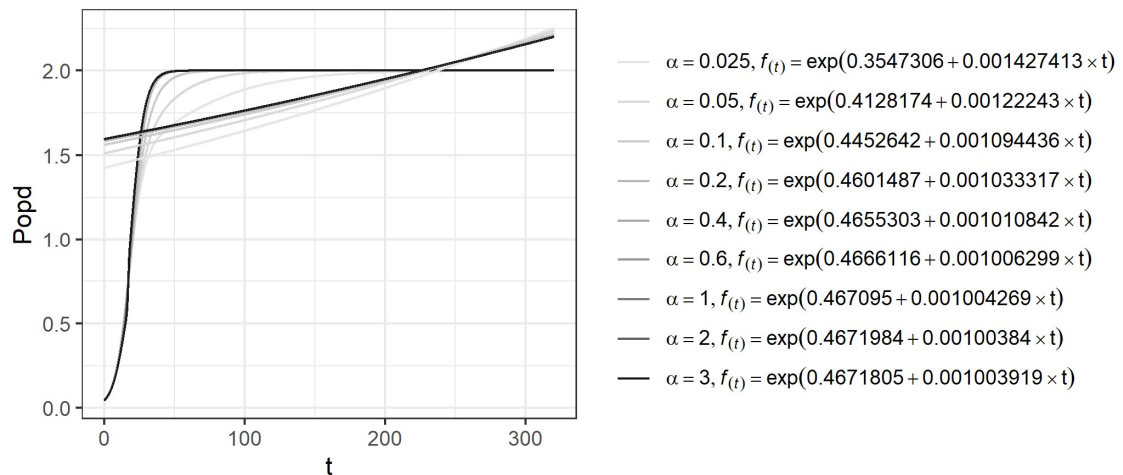


Figure 2.7 Exponential curves fit to bilogistic growth model for nine values  $\alpha$  and one value  $k_g$ , 320 time steps.

At the greater timescale, the simulated demographic curves reach equilibrium and the anomalous reversal at lower alpha values disappears, but the increase with decreasing  $\alpha$  values remains. From this, we can deduce that if the data truncate before the curve reaches saturation, a regression will be blind to whether the latest value in the data represent a lower  $k_g$  value or an incomplete curve. Second—and predictably—the latest portion of the exponential curve must shift downward to capture non-growth data following a period of growth, such as a population resting at saturation, a population collapse, or edge effects from calibration.

One problem we face in the archaeological record is the radiocarbon curves for any pair of regions will have different histories and may be in different phases of demographic growth. Even in the same phase of growth, they may be at different proximity to equilibrium. This would make an interpretation of the differences in growth coefficients estimated via exponential regression challenging.

When comparing two demographic curves using naïvely fit exponential regressions, the rate of adoption and the amount of change in demographic saturation level interact, and it could be difficult to assess and interpret this interaction in real world terms. The amount of right or left shift indicated by the translation coefficient of the exponential regression could serve as a useful clue as to how quickly a society adopts new strategies. However, the translation coefficient receives little or no treatment in archaeological literature, and its presence in exponential regressions seems more grounded in measures of goodness-of-fit than in reference to theoretical concepts relevant to demography. Furthermore, the presence, timing, and rate of depopulation events, the question of whether or not data truncate before a population reaches equilibrium, and

whether the data capture radiocarbon evidence of a population resting at equilibrium all complicate the interpretability of an exponential curve.

Importantly, the amount of change in the exponential growth rate arising from differential rates of adoption of the new strategy appear also to be secondary to the amount of change in the exponential growth rate arising from differential amounts of growth; holding the rate of adoption constant, the general pattern holds true that increased population saturation levels increase the exponentially regressed growth rate of the demographic curve despite a constant intrinsic growth rate. Unfortunately, holding constant the rate of change in saturation level would require a knowledge of the relationship between multi-logistic curves and the radiocarbon curve—the relationship I explore in this paper.

### **Hypotheses for archaeological radiocarbon.**

(1) It is relative changes in population saturation that matter, not absolute changes. The more that technological and social organization change population saturation any region, the faster an exponential growth rate will appear. In the radiocarbon record, large relative changes in  $k$  would take the appearance of steep growth despite a constant intrinsic rate of population growth.

(2) It is possible that the decisions made when fitting an exponential model to the radiocarbon curve might affect the numeric value of the exponential growth rate. For example, where a researcher trims an SPD to mitigate the effects of calibration may result in a deeper or more shallow curve, depending on where in a logistic growth curve the data truncates.

## CHAPTER 3 DATA AND METHODS

I ran exponential and multilogistic regression models on radiocarbon records from four regions where hunter-gatherer populations persisted throughout the Holocene in N. America. For each region, I assembled SPDs from calibrated radiocarbon dates. I predicted that the regions with larger relative differences in  $k$  parameters ( $k_g$ ) would have higher long-term growth rates, as estimated by exponential and multilogistic regression. I then placed the regions in rank-order based on these growth rates. If an exponential regression adequately describes the multi-stage logistic growth processes encapsulated in the Malthus-Boserup ratcheting model, a rank-ordering of regions based on exponential growth rates should match the rank-ordered list of regions based on  $k_g$ .

Alternatively, the descriptions provided by multilogistic and exponential regressions could be fundamentally incompatible. Finally, I explore information loss associated with both regression models using the Bayesian Information Criterion for small sample size (BIC). I use BIC for two reasons: first, as a second-order calculation, it is more reliable than Akaike Information Criterion; second, it penalizes models for using more parameters, taking into account sample size effects on the residual sum of squares—where Akaike Information Criterion with correction for small sample size (AICc) penalizes parameters by a factor of 2, BIC penalizes parameters by a factor of  $\log(n)$ , resulting in more parsimonious selection.

### **Methods**

Because anthropogenic radiocarbon represents the material remains of human activity, I assumed that changes in anthropogenic radiocarbon deposits over time



represent relative changes in intensity of human habitation (Berry, 1981; Chaput et al., 2015; French et al., 2021; Rick, 1987; Williams, 2012). Under this assumption, and because radiocarbon can be used for dating materials as old as approximately 50,000 years (Walker, 2005), radiocarbon should serve as a reasonable proxy for demographic data.

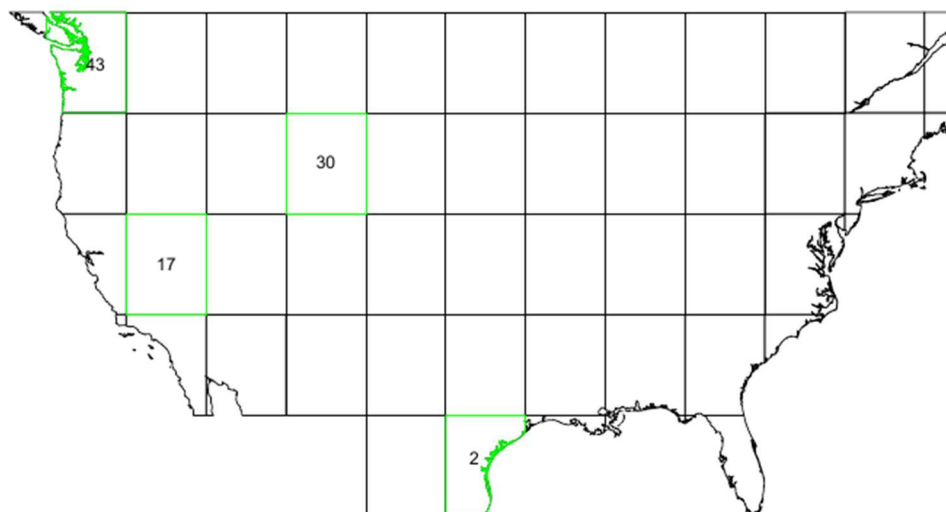
The use of radiocarbon as a proxy for demographic data is not without complications, however; namely, sampling error (Armit et al., 2013; Williams, 2012), calibration effects (Williams, 2012), taphonomic bias (Bluhm and Surovell, 2019; Surovell et al., 2009; Surovell and Brantingham, 2007; Williams, 2012), and the likelihood that the production of anthropogenic carbon deposits scale sub linearly with population among prehistoric people in much the same way as contemporary energy consumption/population relationships (Freeman et al., 2018b). The last of these potential complications means that an adjustment may be necessary if I were attempting to produce actual, rather than relative, measures of demographic change.

My aim in this study is to test the correspondence of exponential growth models, that have very little theoretical underpinning, and multilogistic models that operationalize Malthusian and Boserupian processes underlying demographic changes that may exist in radiocarbon curves. I have left this complication untreated, and suggest that rather than representing changes in demographic saturation itself,  $k_g$  values identified in this study represent changes in the level of waste that people can generate because of changes in demographic saturation. Essentially, for this study,  $k$  is to demographic saturation level as radiocarbon is to population.

I gathered radiocarbon data for North America from published sources and

databases (Bird et al., 2020). To ensure the quality of my data, I took the following steps: 1) I removed dates with duplicated lab numbers. 2) In cases where the data agreed and represented a true duplicate, I only removed one, and in cases where the data disagreed, I opted to remove all radiocarbon dates with that lab number, rather than arbitrarily selecting one and potentially increasing sampling error. 3) I removed records missing date, standard deviation, or complete site number. 4) I removed dates for which the listed material indicated geological or paleontological origins, aggregated samples, or for which the listed material was “unknown,” or left blank, as it is unclear whether these materials are archaeological, geological, paleontological, or aggregated. 5) I removed dates without both latitude and longitude, as geographic location is relevant in this study.

I assigned each date a sampling unit number based on its geographical location within a 5-degree grid that covers the US and Canada (obtained from Bird et al., 2020). I selected two units with coastal areas and two with only inland areas: Texas (including coastal plains and inland), Pacific Northwest, Southwest Wyoming, and Nevada (Figure 3.1).



*Figure 3.1 Sampling units selected for this study*

To ensure that undersampling does not affect my results, I checked that the sample size was reasonable given the time depth of my study and mean laboratory error for each unit (Michczyńska and Pazdur, 2004). This required no complicated steps—the time depth of this study is less than 14,000 years, and the mean uncertainty for each unit is less than 115 years, so a 200-date minimum is adequate (see Michczyńska and Pazdur, 2004). Additionally, the least-populated unit contains 915 dates, which exceeds Michczyńska and Pazdur’s (2004) recommended minimum, as well as Williams’ (2012) recommended minimum of 200-500.

Because the proportion of environmental radiocarbon fluctuates (Bronk Ramsey, 2008), understanding the probability distribution of radiocarbon dating requires calibration of raw radiocarbon dates. I calibrated the dates to the intcal20 (Reimer et al., 2020) calibration curve using the “calibrate” function from the rcarbon package in R

(Bronk Ramsey, 2008; Crema and Bevan, 2021). The output of the calibrate function is the probability distribution of each record.

Summing the probability distributions of these records can produce spikes and troughs resulting from intersite variability in sample size within each region (sample bias). If sample size correlates with population within a region, any thinning of dates artificially decreases the calculated probability of finding anthropogenic radiocarbon at sampling units within that region during periods of more clustered habitation, which potentially decreases our relative measure of population (Bevan and Crema, 2021). Similarly, weighting probabilities based on apparent population at sampling units within a region underestimates population at periods of higher dispersion (Bevan and Crema, 2021). For these reasons, I have chosen to leave sample bias untreated, neither trimming nor binning radiocarbon data.

I generated non-normalized SPDs using the spd function with a running mean of 250 years. To control for edge effects, I generated SPDs for the date range from 12000 to 0 BP, and truncated my study at 550 BP. For this study, I assumed that the number of records overlapping a given year truly represent the relative likelihood of finding anthropogenic carbon materials dating to that year, but I acknowledge that noise in the radiocarbon curve may be more representative of atmospheric carbon than changes in demography. Accordingly, I focused on statistical descriptions of the long-term changes in radiocarbon records rather than interpreting short-term “wiggles.”

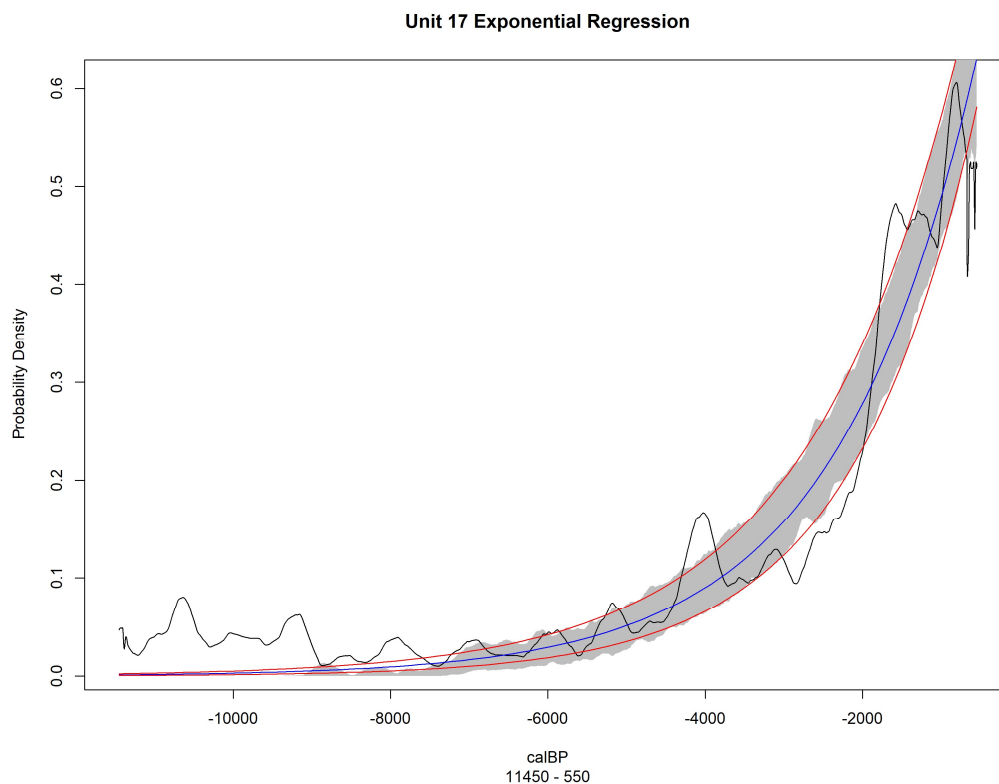
Taphonomic bias is the erasure of materials from their sedimentary context through natural processes. The longer archaeological materials remain exposed to destructive natural processes, the greater their erasure from the archaeological record. As

a result, younger materials, including radiocarbon, may be more highly represented than older materials (Bluhm and Surovell, 2019; Surovell et al., 2009; Surovell and Brantingham, 2007). This means that an SPD will naturally have a higher slope than the actual frequency of material deposition over time (but see Freeman et al., 2018b). Bluhm and Surovell (2019) demonstrate that the effects of taphonomic bias are approximately the same globally. To me, this suggests that applying Bluhm and Surovell's (2019) global taphonomic correction to each dataset would change the slopes, but not rank order of regressions taken on those slopes. Further, Bluhm and Surovell (2019) recommend against using the global taphonomic correction on data later than about 2000 BP—the tail end of the data herein analyzed. For these reasons, I have opted to acknowledge that the slopes are all similarly affected by taphonomic bias, but have left the data uncorrected.

Regressions are a statistical method for finding the best-fit values for parameters of mathematical models. Exponential models, and by extension an exponential regressions, are often used to describe the growth of SPDs in the archaeological literature (Bird et al., 2020; Broughton and Weitzel, 2018; Brown, 2015; Freeman et al., 2021; Shennan et al., 2013; Timpson et al., 2014). Similarly, a logistic regression finds the best-fit values for parameters in a logistic model. A logistic model is appropriate for data where the growth initially appears exponential, but the rate of growth decreases as the measured variables approach some upper boundary, as would be apparent in a plot of saturation levels with an increase. A multilogistic regression finds the best-fit values for parameters in a model that mathematically combines multiple logistic expressions, as would be apparent in a plot of saturation levels with multiple changes. I performed this analysis using exponential and multilogistic regressions without detrending or linearizing

the data.

I ran exponential regressions using R function “nls” from the “stats” package, and stored the translation coefficient and exponential growth rates in a data frame. For this study, statistically distinct growth rates would be ideal. To determine whether the slopes and exponential growth rates were statistically distinct, I checked the reported growth rate against the range of growth rates that would describe 95% of the data. Obtaining the correct confidence interval was slightly more complex than I had anticipated. First, confidence intervals for regression coefficients rely on standard error (SE). Because the sample size is in the denominator in the SE calculation, the sample size of ~12,000 years drives down the standard error. The standard deviation of each point does not face this same challenge. Second, given the somewhat fine-grained nature of SPDs, each wiggle is represented in the data as a year-to-year run. As a result, I suspect that groups of neighboring points differ from the model more similarly than more distant points, producing positive autocorrelation in the residuals. Autocorrelated residuals artificially shrink confidence intervals (Kutner, 2005). To avoid working with artificially small confidence intervals, I obtained the confidence envelopes for regression coefficients by first running the `modelTest` function—which can be used to produce a point-by-point confidence envelope—from the `rcarbon` package, with 1000 iterations. I then extracted the confidence envelopes and ran an exponential regression to succinctly describe them.



*Figure 3.2 SPD for Nevada (Unit 17) plotted over confidence envelope for exponential regression*

I could find no consensus on the best or most valid way of deciding the number of logistic curves to include, as applying multi-logistic waveform analysis to demographic growth curves is relatively unexplored. Freeman et al. (2021) reason that radiocarbon time series with more than one demographic saturation should “display structural breaks in their means, *independent of time*, that indicate multiple levels of population saturation.” They test this by checking for clustering in the relationship between the summed probability of radiocarbon ages and SPD ‘per capita’ growth rate.

I calculated the SPD ‘per capita’ growth rate as

$$g_i = \log(p_i/p_{i-1})$$

## (3.1)

and tested the relationship between  $PrDens$  and  $g_i$  for clustering using the `Hopkins.index` function in the `comato` package in R, which returns the Hopkins index. The Hopkins index is calculated by creating a dataset by uniformly sampling random points from the dataset being tested and dividing the sum of distances between points in the resultant dataset by the aggregated sums of distances between points in each dataset (original and sample; Han et al., 2012)). If there are no meaningful clusters, the real data should be uniformly distributed, and the distances of both datasets should be similar, resulting in a Hopkins index around 0.5 (Han et al., 2012). The more tightly clustered the real data, the greater the effect of the sum of distances of the artificial data in both the numerator and denominator, resulting in a Hopkins Index closer to 1 (Han et al., 2012).

Because the Hopkins Index relies on randomization, it is possible—however unlikely—that any single Hopkins test would report a high likelihood of clustering where there is no clustering, or a low likelihood of clustering where clustering exists. To improve the likelihood that the Hopkins index appropriately discriminates the presence or absence of clusters, and in the interest of reproducibility, I set the seed of R’s random number generator to 2173043 and ran the Hopkins test 25 times, reporting the mean and 95% confidence interval to get an idea of the degree of clustering in the relationship between probability density and growth rate (Table 3.1). The Hopkins Indices for all datasets in this study were greater than 0.9. I interpret this to mean that the data have two or more distinct clusters.

*Table 3.1 Hopkins Indices report degree of clustering, ranging from 0.5 (no clustering) to*



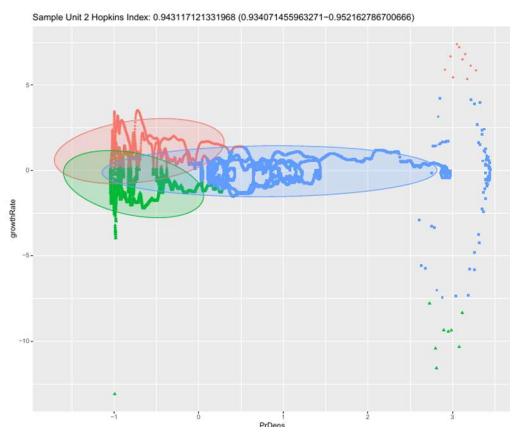
&lt; 1

	<i>HopkinsIndex</i>
Texas	0.9431 (0.9341 – 0.9522)
Nevada	0.9989 (0.9986 – 0.9991)
Wyoming	0.9994 (0.9993 – 0.9995)
PNW	0.9989 (0.9979 – 0.9999)

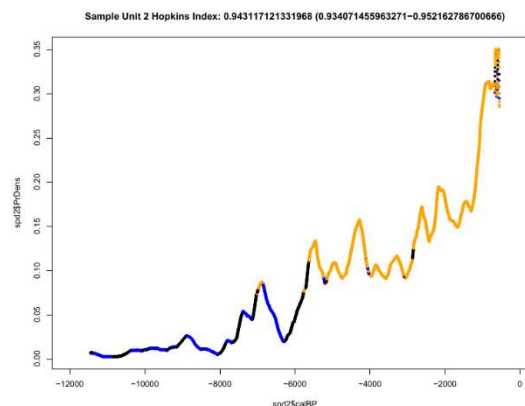
Knowing the number of clusters is important because each cluster potentially represents a unique population saturation level (Freeman et al. 2021). A seemingly straightforward method for estimating the number of unique saturation levels to include in a regression analysis would be to count changes in material culture for each region. However, because of the apparent randomness in SPDs, this method could result in overfitting: some changes in saturation may not alter the parameters of a regression in a meaningful way; a regression with fewer parameters may not, in these cases, result in a loss of the model's descriptive abilities. Simply put, some changes in material culture may signal a lateral change or no change in saturation level. I would need a test to see which of these changes in material culture changed the saturation level, and which did not. The focus of this paper, however is not to test whether changes in material culture correspond with potential demographic inflections in radiocarbon curves.

I simply count the number of clusters that may be influencing the Hopkins Index. There are many methods for counting clusters, and data scientists have yet to agree on

what would be the best method. For each dataset, I plotted each point's growth rate by its probability density to develop an intuitive expectation for the number of clusters. All datasets appeared to have between two and four clusters. I then used the pamk function from the fpc package in R to estimate the number of clusters more concretely in each dataset, restricting the search to between two and four clusters. I used fviz\_cluster from the package factoextra to both visualize clusters and assign a numeric cluster name to each point in the data set. I then plotted each radiocarbon curve, coloring points based on the cluster into which the pamk function had assigned them.



*Figure 3.3 Cluster Plot with ellipsoids for Texas (Unit 2)*



*Figure 3.4 SPD for Texas (Unit 2), colored by cluster*

Inspecting these plots, pamk detected two clusters in two datasets and more than two clusters in two datasets. In the Texas data, two clusters described growth, while a third described downward slopes on “wiggles (Figure 3.4).” For this analysis, I dismissed these downward slopes on “wiggles” as spuriously clustered, and regressed the data on a bilogistic curve. In the PNW dataset, there were three distinct clusters (Figures 3.5 and

3.6). I have provided the plots for all regions in Appendix A.

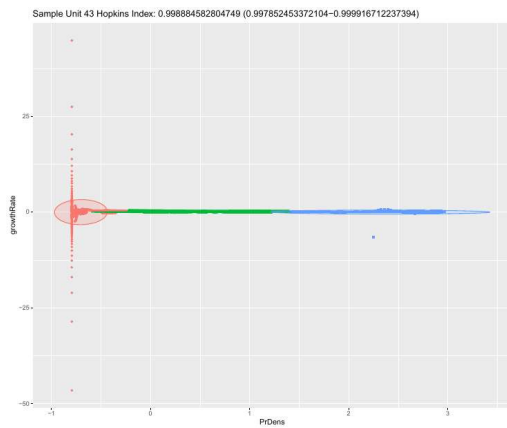


Figure 3.5 Cluster Plot with ellipsoids for  
PNW (Unit 43)

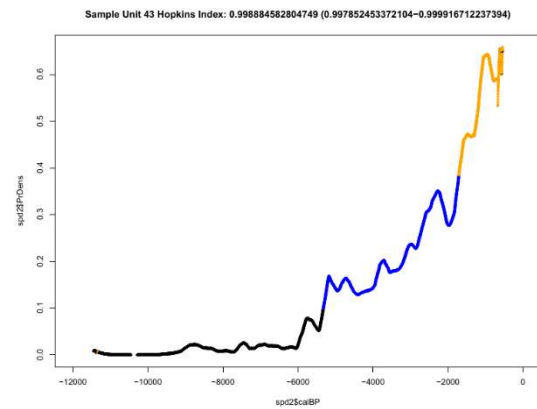


Figure 3.6 SPD for PNW (Unit 43),  
colored by cluster

The `drm` function in the R package `drc` was created for dose-response modeling, a necessary step for describing dose-response relationships in dose-response analyses in the field of toxicology. As with other regression functions, `drm` allows end-users to specify models, but the package includes functions that specifies models and finds appropriate starting values for parameters in the regression. Among the functions for specifying models and finding appropriate starting values are `L.4` and `L.3`. The former refers to the standard logistic curve of the form

$$Y = d + \frac{k-d}{1+e^{-\alpha(t-t_{pi})}} \quad (3.2)$$

where  $k$  is the upper asymptote,  $d$  is the lower asymptote,  $\alpha$  is the intrinsic growth rate,  $t$  is time and  $t_{pi}$  represents the time at the curve's inflection. The latter function forces the lower asymptote,  $c$ , to zero, leaving the model with 3 parameters. For the first portion of each multilogistic curve, I ran `drm` with `L.4` on the subset of the data from the earliest to

1000 years following the latest appearance of data in what substantially appears to be the first cluster. On interior clusters, I began the subset of data 1000 years earlier than the earliest apparent break in the cluster and ended it 1000 years later than the next apparent break in clusters (relevant in this study only to Unit 43), subtracted the effects of earlier prediction curves, and ran `drm` with L.3. For the latest clusters, I similarly began the subset of data 1000 years earlier than the apparent final break in clusters and ended it at zero, and then subtracted the summed effects of earlier prediction curves (Meyer et al, 1999).

The `drm` function accepts a parameter specifying the user-preferred method for robust estimation. Available robust estimation methods include ‘mean’ (the default method, which is non-robust), ‘median,’ least median of squares (‘lms’), and least trimmed squares (‘lts’). Through experimentation, I found that the goodness of fit of regressions taken on each subset of data might be affected differently from other subsets depending on the rho function used. For example, subset 1 might achieve the best fit with the ‘median’ method of estimation, while subset 2 of the same SPD might achieve best fit with the ‘lms’ method. I ran the `drm` function 4 times for each subset, once for each method of estimation, and then assembled each permutation of summed models representing each SPD in its entirety, leaving the lower asymptote unsolved.

I left the lower asymptote unsolved because there is no function in R for calculating a single BIC on three separate `drc` objects, so I would need to run the summed model through the `nls` function to generate an `nls` object. To avoid `nls` altering the `d` parameter as discovered through `drm`, I specified the parameter’s starting value, as well as its upper and lower bounds, in `nls` with the value discovered using `drm`. I then

extracted the BIC for each nls object representing each permutation of estimation methods and kept the model arising from the permutation with the least information loss. Because BIC relies on the number of  $\beta$  parameters reported by the nls object, the BIC value reported for the nls object is too low by  $\log(n) * 3$  for each cluster. To remedy this, I add to the reported BIC value  $\log(n) * 3x$  where  $x$  represents the number of clusters.

*Table 3.2 Best fit estimation used for each wave*

Best Fit Permutation of Estimation Methods			
	1 <sup>st</sup> Wave	2 <sup>nd</sup> Wave	3 <sup>rd</sup> Wave
Unit 2 (Texas)	“lms”	“mean”	NA
Unit 17 (Nevada)	“lms”	“lts”	NA
Unit 30 (Wyoming)	“lts”	“mean”	NA
Unit 43 (PNW)	“mean”	“lts”	“mean”

I then extracted the upper asymptotes from each regression taken on each subset of data, subtracted the lower asymptote if present, and calculated  $k_g$  as the relative change in  $k$  for that SPD. Specifically,

$$k_g = (\sum_{i=2}^n k_i) / (k|1 - d). \quad (3.3)$$

Lastly, I compared the information loss associated with each model for the four selected regions by running the BIC function in R on the exponential models. Bilogistic and multilogistic models describe curves in segments, each segment alone being more

complex than the standard exponential model. BIC is appropriate for this study because it measures information loss with a penalty for model complexity. Unlike AICc, however, BIC accounts for the effect sample size has on RSS by scaling the penalty with the natural log of the sample size

## CHAPTER 4 EMPIRICAL RESULTS

I set out to explore the correspondence of population growth described by exponential curves and the processes of demographic change arising from Malthus-Boserup ratcheting processes that may generate differences in long-term growth rates of SPDs. The assumption I have tested is simple: if the exponential curve adequately describes demographic change arising from Malthus-Boserup ratcheting, then the growth rate of an exponential regression taken on SPDs should be higher with higher relative changes in saturation level, and lower with lower relative changes in saturation level. To test this, I took exponential regressions on SPDs, ranked the SPDs from highest to lowest of each of these coefficients, and compared these rank-ordered SPDs to SPDs ranked by relative change in saturation level as calculated from multiple logistic regressions. If the rank orders from the regressions match the rank order from the waveform analysis, relative changes in saturation predict growth rates.

Table 4.1 shows the exponential growth rate as extracted from each exponential regression, along with the exponential growth rates from exponential regressions taken on the high and low boundaries of the confidence envelopes.

*Table 4.1 Exponential growth rates with the growth rates of upper and lower bounds of confidence envelope*

	<i>GrowthRate</i>
Texas (Unit 2)	0.000296 (0.000258-0.000359)
Nevada (Unit 17)	0.000565 (0.000523-0.000632)

Wyoming (Unit 30)	0.000395 (0.000371-0.000443)
PNW (Unit 43)	0.000446 (0.000413-0.000504)

The confidence envelopes for Wyoming and the Pacific Northwest overlap, but the reported exponential growth rate for each of the four units falls within its own confidence envelope and no other. To me, this would suggest the possibility that the true exponential growth rates for Wyoming and the Pacific Northwest may have the same long-term growth rates. Table 4.2 shows the rank order of growth rates estimated via exponential regression.

*Table 4.2 Rank order of regions based on exponential growth*

Predicted order by exponential growth (Highest to Lowest)
Nevada
PNW
Wyoming
Texas

Table 4.3 shows the results of the multilogistic regressions. Note that the upper asymptote of the first logistic curve subtracts the lower asymptote. I have left this unsimplified so as to avoid confusion surrounding the calculation of relative growth in saturation.

*Table 4.3 Multilogistic regression results.*

Summed Logistic Curves
------------------------



Texas (Unit 2)	$\left(\frac{0.1324}{1 + e^{0.0009(-6197 - calBP)}}\right)$ $+ \left(\frac{0.3685}{1 + e^{0.0039*(-913 - calBP)}}\right) + 0.0023$
Nevada (Unit 17)	$\left(\frac{0.5952}{1 + e^{0.0006(-508 - calBP)}}\right)$ $+ \left(\frac{0.1836}{1 + e^{0.0126(-1904 - calBP)}}\right) + 0.0099$
Wyoming (Unit 30)	$\left(\frac{0.2090}{1 + e^{0.0005(-4873 - calBP)}}\right)$ $+ \left(\frac{0.8202}{1 + e^{0.0102(-1741 - calBP)}}\right) + 0.0144$
PNW (Unit 43)	$\left(\frac{0.1436}{1 + e^{0.0040(-5547 - calBP)}}\right)$ $+ \left(\frac{0.1308}{1 + e^{0.0040(-3038 - calBP)}}\right)$ $+ \left(\frac{0.3572}{1 + e^{0.0040(-1462 - calBP)}}\right) + 0.0090$

The numerator in the first logistic curve represents the initial upper asymptote—the mathematical equivalent of the saturation level from 11450 BP to the change in saturation. In other words, using the earlier strategy in Nevada could only support a population capable of producing enough anthropogenic radiocarbon that with contemporary archaeological methods, we would only detect a maximum probability density of 0.5952 (the numerator in the first term). People would begin feeling the effects of demographic pressure at about 0.2976 (half the numerator in the first term), and the change in strategy increased the saturation level by about 31% (the second numerator divided by the first numerator).

It is important to note that while the first logistic wave of this equation would indicate that the zone of demographic pressure should start at around 508 cal BP, the value nearest to 50% of this early saturation in the SPD appears at 1888 cal BP. I found this value by looking up the date associated with the minimum absolute value of the difference between PrDens and 0.2976. This would suggest that the regression produced low estimates at the later end of the first subset of data. The correction for this underestimation appears in the second logistic wave, where the inflection appears at 1904 cal BP, earlier than both the previous inflection point and the looked-up value of 1888. This would suggest one of the following possibilities: 1) we are not seeing Malthus-Boserup ratcheting in Nevada; 2) I set the upper limit on the number of clusters too low; 3) noise in the radiocarbon data resulted in inaccurate cluster detection; or 4) the change in strategy only provided partial relief from the effects of demographic pressure.

Table (4.4) shows the relative growth in saturation as calculated from the multilogistic regressions.

*Table 4.4 Relative change in saturation level as calculated from multilogistic regression*

	$\frac{k_{g(total)}}{\left(\frac{\sum_{i=2}^n k_i}{(k_1 - d)}\right)}$	$\left(\frac{k_{g1}}{k_2}\right)$	$\left(\frac{k_{g2}}{k_3}\right)$
Texas	2.785	2.785	NA
Nevada	0.308	0.308	NA
Wyoming	3.925	3.925	NA
PNW	3.398	0.911	1.302

Table 4.5 shows the observed order of regions as ranked by relative growth in saturation level.

Table 4.5 Regions rank ordered by relative growth in saturation

Observed order by total relative growth in saturation (Highest to Lowest)	Observed order by relative growth in saturation, first $k_g$ (Highest to Lowest)	Observed order by relative growth in saturation, last $k_g$ (Highest to Lowest)
Wyoming	Wyoming	Wyoming
PNW	Texas	Texas
Texas	PNW	PNW
Nevada	Nevada	Nevada

The order of regions ranked by relative growth does not match the order of regions ranked by exponential growth rate. This means that differences in growth rate do not vary as I would suspect based on changes in population saturation noted in Chapter 2.

Finally, Table 4.6 shows a side-by-side comparison of calculated BIC for each model. The amount of information conserved in the multilogistic model justifies its use, despite the heavier penalty for complexity imposed by the BIC calculation. A BIC difference between 0 and 2 is interpreted as providing “weak” evidence in favor of the model with the lower BIC value. A difference between 2 and 6 provides “positive” evidence; between 6 and 10, “strong” evidence, and greater than 10 provides “very strong” evidence (Raftery, 1995). In all cases, differences in BIC provide “very strong” support for use of the multilogistic model over the exponential.

Table 4.6 BIC values for exponential and multilogistic regressions, and the difference in BIC values

BIC			
	Exponential	Multilogistic	$\Delta BIC$
Texas	-48542	-51216	2674
Nevada	-38024	-43682	5658
Wyoming	-3857	-8962	5105
PNW	-46716	-54900	8184

## CHAPTER 5 DISCUSSION AND CONCLUSION

**Discussion**

In this thesis, I have attempted to make progress on answering the following question: *Do hunter-gatherers experience the same long-term trajectories of population growth and growth rates regardless of ecological context?* Answering this question is difficult. One approach is to use archaeological radiocarbon, construct summed probability distributions, and then statistically fit a population growth model to summed probability distributions. There are several sets of challenges associated with this approach. One set may be called bias challenges: Archaeologists face sample bias and taphonomic bias in the collection of radiocarbon ages and calibration biases when calibrating ages. A second set of challenges, I call modeling challenges, receive less attention. Modeling challenges refer to the process of choosing an appropriate population growth model to fit to SPDs and interpreting the coefficients of the model. This has been, in part, the focus of my thesis. In order to compare population growth rates across regions, all else equal, we must choose an appropriate model and understand how to interpret the model's parameters. To gain insight into this, I explored two different population growth models in this thesis: The exponential and a summed multilogistic model.

The exponential model is most commonly used in N. America to describe the growth of radiocarbon records, SPDs in particular. However, it is not clear why humans would be the only animal that experiences exponential-like growth rather than density dependent growth. One possibility is that humans can innovate and, in an emergent way, increase their carrying capacity. To capture this possibility, I proposed a bi-logistic model

of population growth to operationalize the so called Malthus-Boserup Ratchet theory of human population expansion (Wood, 1998). I ran simulations of the bilogistic model to better understand the model's behavior and to understand the kind of biases that might arise if one fit an exponential model to a population that was actually undergoing multiple components of density dependent growth. This exercise was technical, but very important. If we are to compare the growth rates of populations across regions using models, we must understand the underlying behavior of the models in order to interpret their parameters. This simulation exercise taught me that if populations are growing according to some complex set of density dependent processes, then it is very difficult, if not impossible to compare the parameters of an exponential model across regions. Multiple factors, including the magnitude of change in carrying capacity and rates of innovation, may drive variation in estimated growth rates from an exponential model.

I, thus, concluded that it might be better to fit multilogistic models to SPDs than the exponential. In the empirical portion of my thesis, I explored whether rank ordering SPDs by growth rate from exponential regressions predicted the order of SPDs ranked by relative changes in population saturation level calculated from multilogistic regression. If the observed pattern of demographic change across N. America was ultimately the same from region to region, it might have been reasonable to suspect that the rank orders derived from exponential and multilogistic regressions would match, given the models under which I developed my predictions. However, from the multilogistic regression, we see that the timing and relative size of changes in saturation tell the story of a unique demographic history within each region. This interpretive data is neither encapsulated in the exponential growth rate nor the exponential translation coefficient.

Ultimately, I argue that fitting multilogistic models gives us potentially more information and the ability to generate hypotheses about other aspects of the archaeological record across regions. For example, along the NW Coast of North America, we observe a population growth component from approximately 8000 to 4000 cal BP. If this growth component is useful, we should be able to use it to predict other changes in archaeological records. In particular, we would expect evidence in faunal records of more use of high ranked resources at the beginning of the growth component 8000 to 6000 cal BP and a shift to lower ranked faunal resources 5999 to 4000 cal BP as population pressure began to bite on foragers by decreasing the encounter rates of larger game. Similarly, the multilogistic model would suggest a shift back toward higher ranked resources after 4000 cal BP as a new growth component began.

Although the multilogistic models are based on a logically coherent theory (the Malthus-Boserup Ratchet), there are obviously still all the bias challenges associated with using these models to describe radiocarbon curves. Here are several that arose during my study:

(1) Autocorrelated Residuals: I found autocorrelation in the residuals of both the exponential and logistic models. This makes sense, given that SPDs represent a trend in radiocarbon deposition over an interval of time. Put simply, this means that as the radiocarbon curve departs from the model toward local peaks and valleys, it does so at a measurable slope (though perhaps with some error), and we can predict this departure at time  $t$  using the error at  $t - 1$ . This violates the assumption that observations are independent of one another. Procedures for remedying this violation are aimed at detrending the data, but our aim is to describe the trend, not the data's most likely

relationship to time as a predictor (as there is none). One possible remedy may be to include a lag term in the model so that error at  $t - 1$  is included in both exponential and logistic models.

(2) The BIC estimation relies on the assumption that residuals are not autocorrelated, and much of the magnitude of the  $\Delta\text{BIC}$  values may reflect the degree of autocorrelation in the residuals more than the appropriateness of the models themselves. On the other hand, autocorrelation in the residuals can be interpreted as a reflection of the degree to which a model is mis-specified. My application of BIC in this case is debatable, as would be AIC or AICc. What is important is the recognition that an exponential model violates the assumption of density dependence in population growth, and that in the archaeological discipline, we need to begin addressing modeling challenges.

(3) I have assumed (following Freeman 2021) that the correct number of logistic models to include correlates directly with the number of clusters observed in a plot of growth rates against probability. While the rationale makes sense, I think we need to explore this more in noisy data. In a “perfect” single logistic curve, the growth rate would increase from zero with population until the inflection point, where it would peak, and then decrease back to zero as the population stabilized at saturation. Adding a second logistic curve, we should have two such peaks. The problem we encounter in noisy data is that the point-to-point growth rate between any three points does not necessarily follow the same rule. Unsmoothed SPDs have many peaks and valleys, and the points between these peaks and valleys may follow their own line or curve at annual intervals, and we may observe as many peaks in the slope-population relationship as there are peak-valley pairs. The next problem we encounter with this method is that it the cluster detection

algorithm may be clustering based on nothing more than similarity in probability. We saw in the Texas unit that negative slopes were detected as distinct clusters despite the actual steepness of these slopes varying as widely as positive slopes. Further, while I have assumed that the points identified as belonging to a cluster may fall anywhere in time (following Freeman, 2021), PNW, Wyoming, and Nevada all seem to have clustering that divides the vertical distribution of probabilities somewhat evenly. This is particularly evident in Wyoming, where the late part of the SPD dips below the probability identified as a break in clusters. To address this third concern, it may be sensible to nonparametrically smooth SPDs, and then look for what would appear to be significant humps or peaks in the second difference.

## **Conclusion**

Relative changes in an environment's population saturation level resulting from a population's response to demographic pressure may result in exponential-like curves observed in the radiocarbon record. Unfortunately, the exponential growth rate also varies according to decisions made in preparing the data for analysis as well as variations in demographic history. This renders exponential growth rates difficult to compare between regions. The Malthus-Boserup Ratcheting Model provides some potential context by which we can begin to understand and explain regional differences in patterns of demographic growth. My test of the implications of Malthus-Boserup Ratcheting for exponential regressions provides support for the use of multiple logistic regressions in understanding the demographic histories underlying the shape of SPD curves. Moving forward, applying population-ratchet like models can aid in understanding the demographic and socioecological context of cultural changes—a



crucial understanding in the pursuit of the study of the lifeways of past peoples.

## References

- Armit, I., Graeme T. Swindles, Katharina Becker, 2013. From dates to demography in later prehistoric Ireland? Experimental approaches to the meta-analysis of large 14C data-sets. *Journal of Archaeological Science* 40, 433–438.  
<https://doi.org/10.1016/j.jas.2012.08.039>
- Berry, M.S., 1981. *Time, Space, and Transition in Anasazi Prehistory*.
- Biermann, F., Abbott, K., Andresen, S., Backstrand, K., Bernstein, S., Betsill, M.M., Bulkeley, H., Cashore, B., Clapp, J., Folke, C., Gupta, A., Gupta, J., Haas, P.M., Jordan, A., Kanie, N., Kluvankova-Oravska, T., Lebel, L., Liverman, D., Meadowcroft, J., Mitchell, R.B., Newell, P., Oberthur, S., Olsson, L., Pattberg, P., Sanchez-Rodriguez, R., Schroeder, H., Underdal, A., Vieira, S.C., Vogel, C., Young, O.R., Brock, A., Zondervan, R., 2012. Navigating the Anthropocene: Improving Earth System Governance. *Science* 335, 1306–1307.  
<https://doi.org/10.1126/science.1217255>
- Binford, L.R., 2002. *In pursuit of the past: decoding the archaeological record: with a new afterword*, 1st Calif. pbk. ed. University of California Press, Berkeley.
- Bird, D., Freeman, J., Robinson, E., Maughan, G., Finley, J.B., Lambert, P.M., Kelly, R.L., 2020. A first empirical analysis of population stability in North America using radiocarbon records. *The Holocene* 30, 1345–1359.  
<https://doi.org/10.1177/0959683620919975>
- Bluhm, L.E., Surovell, T.A., 2019. Validation of a global model of taphonomic bias using geologic radiocarbon ages. *Quat. res.* 91, 325–328.  
<https://doi.org/10.1017/qua.2018.78>
- Bronk Ramsey, C., 2008. RADIOCARBON DATING: REVOLUTIONS IN UNDERSTANDING. *Archaeometry* 50, 249–275. <https://doi.org/10.1111/j.1475-4754.2008.00394.x>
- Broughton, J.M., Weitzel, E.M., 2018. Population reconstructions for humans and megafauna suggest mixed causes for North American Pleistocene extinctions. *Nat Commun* 9, 5441. <https://doi.org/10.1038/s41467-018-07897-1>
- Brown, W.A., 2015. Through a filter, darkly: population size estimation, systematic error, and random error in radiocarbon-supported demographic temporal frequency analysis. *Journal of Archaeological Science* 53, 133–147.  
<https://doi.org/10.1016/j.jas.2014.10.013>
- Catton, W.R., 1987. The World's Most Polymorphic Species. *BioScience* 37, 413–419.  
<https://doi.org/10.2307/1310565>
- Chaput, M.A., Gajewski, K., 2016. Radiocarbon dates as estimates of ancient human population size. *Anthropocene* 15, 3–12.  
<https://doi.org/10.1016/j.ancene.2015.10.002>
- Chaput, M.A., Kriesche, B., Betts, M., Martindale, A., Kulik, R., Schmidt, V., Gajewski, K., 2015. Spatiotemporal distribution of Holocene populations in North America. *Proc Natl Acad Sci USA* 112, 12127–12132.  
<https://doi.org/10.1073/pnas.1505657112>
- Cohen, J.E., 1995. Population growth and earth's human carrying capacity. *Science* 269, 341–346.

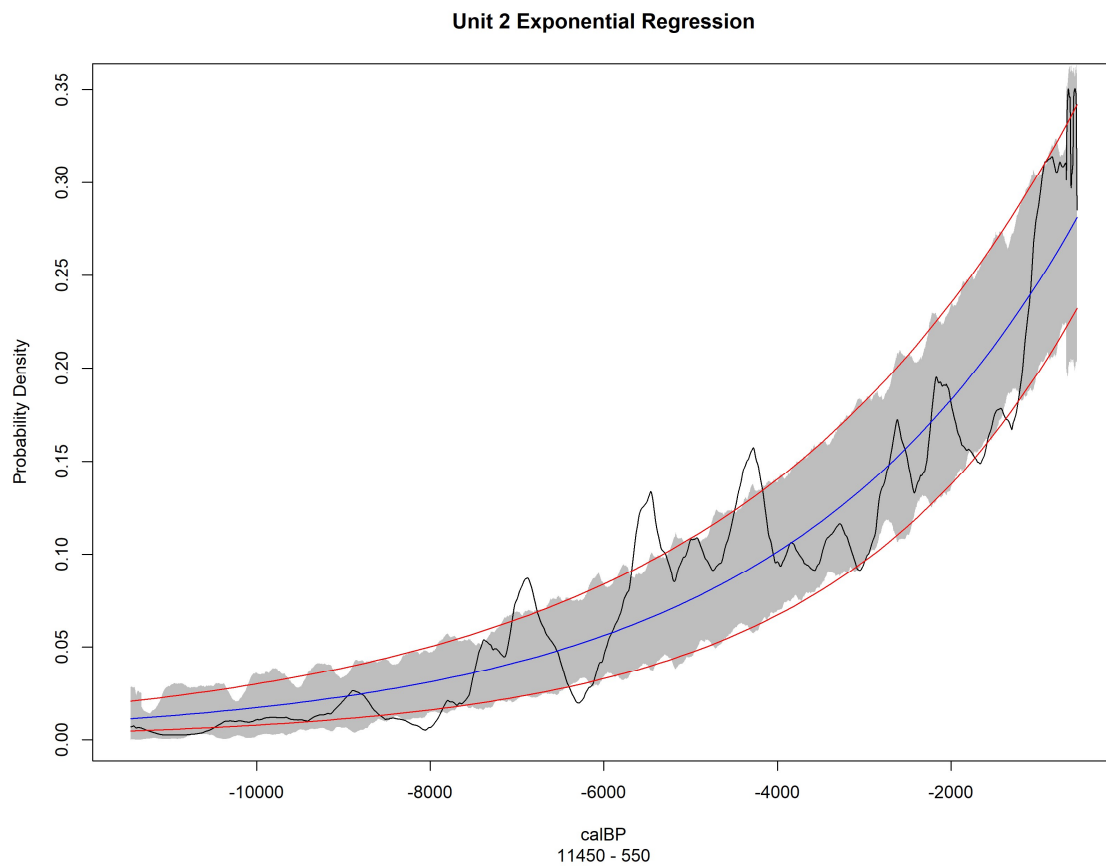
- Crema, E.R., Bevan, A., 2021. Inference from large sets of radiocarbon dates: software and methods. *Radiocarbon* 63, 23–39.
- Dietz, T., Ostrom, E., Stern, P.C., 2003. The Struggle to Govern the Commons 302, 6.
- Fedoroff, N.V., Cohen, J.E., 1999. Plants and population: Is there time? *Proceedings of the National Academy of Sciences* 96, 5903–5907.  
<https://doi.org/10.1073/pnas.96.11.5903>
- Freeman, J., Anderies, J.M., Mauldin, R.P., Hard, R.J., 2019. Should I stay or should I go? The emergence of partitioned land use among human foragers. *PLoS ONE* 14, e0218440. <https://doi.org/10.1371/journal.pone.0218440>
- Freeman, J., Baggio, J.A., Robinson, E., Byers, D.A., Gayo, E., Finley, J.B., Meyer, J.A., Kelly, R.L., Anderies, J.M., 2018a. Synchronization of energy consumption by human societies throughout the Holocene. *Proc Natl Acad Sci USA* 115, 9962–9967. <https://doi.org/10.1073/pnas.1802859115>
- Freeman, J., Byers, D.A., Robinson, E., Kelly, R.L., 2018b. Culture Process and the Interpretation of Radiocarbon Data. *Radiocarbon* 60, 453–467.  
<https://doi.org/10.1017/RDC.2017.124>
- Freeman, J., Hard, R.J., Mauldin, R.P., Anderies, J.M., 2021. Radiocarbon data may support a Malthus-Boserup model of hunter-gatherer population expansion. *Journal of Anthropological Archaeology* 63, 101321.  
<https://doi.org/10.1016/j.jaa.2021.101321>
- Freeman, J., Robinson, E., Beckman, N.G., Bird, D., Baggio, J.A., Anderies, J.M., 2020. The global ecology of human population density and interpreting changes in paleo-population density. *Journal of Archaeological Science* 120, 105168.  
<https://doi.org/10.1016/j.jas.2020.105168>
- French, J.C., Riris, P., Fernández-López de Pablo, J., Lozano, S., Silva, F., 2021. A manifesto for palaeodemography in the twenty-first century. *Phil. Trans. R. Soc. B* 376, 20190707. <https://doi.org/10.1098/rstb.2019.0707>
- Han, J., Kamber, M., Pei, J., 2012. *Data mining: concepts and techniques*, 3rd ed. ed. Elsevier, Burlington, MA.
- Hardin, G., 1968. The Tragedy of the Commons 162, 7.
- Johnson, C.N., Brook, B.W., 2011. Reconstructing the dynamics of ancient human populations from radiocarbon dates: 10 000 years of population growth in Australia. *Proc. R. Soc. B* 278, 3748–3754.  
<https://doi.org/10.1098/rspb.2011.0343>
- Kelly, R.L., Surovell, T.A., Shuman, B.N., Smith, G.M., 2013. A continuous climatic impact on Holocene human population in the Rocky Mountains. *Proceedings of the National Academy of Sciences* 110, 443–447.  
<https://doi.org/10.1073/pnas.1201341110>
- Kutner, M.H. (Ed.), 2005. *Applied linear statistical models*, 5th ed. ed, The McGraw-Hill/Irwin series operations and decision sciences. McGraw-Hill Irwin, Boston.
- Lucarini, G., Wilkinson, T., Crema, E.R., Palombini, A., Bevan, A., Broodbank, C., 2020. The MedAfriCarbon Radiocarbon Database and Web Application. *Archaeological Dynamics in Mediterranean Africa, ca. 9600–700 BC. Journal of Open Archaeology Data* 8, 1. <https://doi.org/10.5334/joad.60>
- Malthus, T., 2003. *An Essay on the Principle of Population*. Project Gutenberg, Urbana, Illinois.

- Martindale, A., Morlan, R., Betts, M., Blake, M., Gajewski, C., Chaput, M., Mason, A., Vermeersch, P., 2016. Canadian Archaeological Radiocarbon Database.
- Meyer, P.S., Ausubel, J.H., 1999. Carrying Capacity: A Model with Logistically Varying Limits. *Technological Forecasting and Social Change* 61, 209–214. [https://doi.org/10.1016/S0040-1625\(99\)00022-0](https://doi.org/10.1016/S0040-1625(99)00022-0)
- Meyer, P.S., Yung, J.W., Ausubel, J.H., 1999. A Primer on Logistic Growth and Substitution. *Technological Forecasting and Social Change* 61, 247–271. [https://doi.org/10.1016/S0040-1625\(99\)00021-9](https://doi.org/10.1016/S0040-1625(99)00021-9)
- Michczyńska, D.J., Pazdur, A., 2004. Shape Analysis of Cumulative Probability Density Function of Radiocarbon Dates Set in the Study of Climate Change in the Late Glacial and Holocene. *Radiocarbon* 46, 733–744. <https://doi.org/10.1017/S0033822200035773>
- Palmisano, A., Bevan, A., Shennan, S., 2018. Regional Demographic Trends and Settlement Patterns in Central Italy: Archaeological Sites and Radiocarbon Dates. *Journal of Open Archaeology Data* 6, 2. <https://doi.org/10.5334/joad.43>
- Pardo-Gordó, S., García Puchol, O., Bernabeu Aubán, J., Díez Castillo, A., 2019. Timing the Mesolithic-Neolithic Transition in the Iberian Peninsula: The Radiocarbon Dataset. *Journal of Open Archaeology Data* 7, 4. <https://doi.org/10.5334/joad.49>
- Peros, M.C., Munoz, S.E., Gajewski, K., Viau, A.E., 2010. Prehistoric demography of North America inferred from radiocarbon data. *Journal of Archaeological Science* 37, 656–664. <https://doi.org/10.1016/j.jas.2009.10.029>
- Puleston, C., Tuljapurkar, S., Winterhalder, B., 2014. The Invisible Cliff: Abrupt Imposition of Malthusian Equilibrium in a Natural-Fertility, Agrarian Society. *PLoS ONE* 9, e87541. <https://doi.org/10.1371/journal.pone.0087541>
- Raftery, A.E., 1995. Bayesian Model Selection in Social Research. *Sociological Methodology* 25, 111. <https://doi.org/10.2307/271063>
- Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., Grootes, P.M., Guilderson, T.P., Hajdas, I., Heaton, T.J., Hogg, A.G., Hughen, K.A., Kromer, B., Manning, S.W., Muscheler, R., Palmer, J.G., Pearson, C., van der Plicht, J., Reimer, R.W., Richards, D.A., Scott, E.M., Southon, J.R., Turney, C.S.M., Wacker, L., Adolphi, F., Büntgen, U., Capano, M., Fahrni, S.M., Fogtmann-Schulz, A., Friedrich, R., Köhler, P., Kudsk, S., Miyake, F., Olsen, J., Reinig, F., Sakamoto, M., Sookdeo, A., Talamo, S., 2020. The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon* 62, 725–757. <https://doi.org/10.1017/RDC.2020.41>
- Rick, J.W., 1987. Dates as Data: An Examination of the Peruvian Preceramic Radiocarbon Record. *American Antiquity* 52, 55. <https://doi.org/10.2307/281060>
- Riede, F., 2009. Climate and Demography in Early Prehistory: Using Calibrated <sup>14</sup>C Dates as Population Proxies. *Human Biology* 81, 309–337. <https://doi.org/10.3378/027.081.0311>
- Robinson, E., Zahid, H.J., Coddling, B.F., Haas, R., Kelly, R.L., 2019. Spatiotemporal dynamics of prehistoric human population growth: Radiocarbon ‘dates as data’ and population ecology models. *Journal of Archaeological Science* 101, 63–71. <https://doi.org/10.1016/j.jas.2018.11.006>
- Rockström, J., Williams, J., Daily, G., Noble, A., Matthews, N., Gordon, L.,

- Wetterstrand, H., DeClerck, F., Shah, M., Steduto, P., de Fraiture, C., Hatibu, N., Unver, O., Bird, J., Sibanda, L., Smith, J., 2017. Sustainable intensification of agriculture for human prosperity and global sustainability. *Ambio* 46, 4–17. <https://doi.org/10.1007/s13280-016-0793-6>
- Shennan, S., Downey, S.S., Timpson, A., Edinborough, K., Colledge, S., Kerig, T., Manning, K., Thomas, M.G., 2013. Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nat Commun* 4, 2486. <https://doi.org/10.1038/ncomms3486>
- Surovell, T.A., Brantingham, P.J., 2007. A note on the use of temporal frequency distributions in studies of prehistoric demography. *Journal of Archaeological Science* 34, 1868–1877. <https://doi.org/10.1016/j.jas.2007.01.003>
- Surovell, T.A., Byrd Finley, J., Smith, G.M., Brantingham, P.J., Kelly, R., 2009. Correcting temporal frequency distributions for taphonomic bias. *Journal of Archaeological Science* 36, 1715–1724. <https://doi.org/10.1016/j.jas.2009.03.029>
- Tallavaara, M., Eronen, J.T., Luoto, M., 2018. Productivity, biodiversity, and pathogens influence the global hunter-gatherer population density. *Proc Natl Acad Sci USA* 115, 1232–1237. <https://doi.org/10.1073/pnas.1715638115>
- Timpson, A., Colledge, S., Crema, E., Edinborough, K., Kerig, T., Manning, K., Thomas, M.G., Shennan, S., 2014. Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. *Journal of Archaeological Science* 52, 549–557. <https://doi.org/10.1016/j.jas.2014.08.011>
- Walker, M.J.C., 2005. *Quaternary dating methods*. J. Wiley, Chichester, West Sussex, England.
- Williams, A.N., 2012. The use of summed radiocarbon probability distributions in archaeology: a review of methods. *Journal of Archaeological Science* 39, 578–589. <https://doi.org/10.1016/j.jas.2011.07.014>
- Wood, J.W., 2020. *The Biodemography of Subsistence Farming: Population, Food and Family*. Cambridge University Press.
- Wood, J.W., 1998. A Theory of Preindustrial Population Dynamics Demography, Economy, and Well-Being in Malthusian Systems. *Current Anthropology* 39, 99–135. <https://doi.org/10.1086/204700>
- Zahid, H.J., Robinson, E., Kelly, R.L., 2016. Agriculture, population growth, and statistical analysis of the radiocarbon record. *Proceedings of the National Academy of Sciences* 113, 931–935. <https://doi.org/10.1073/pnas.1517650112>

APPENDICES

## APPENDIX I: SPDS WITH EXPONENTIAL REGRESSIONS



*Figure AI.1 Exponential regressions for Texas (Unit 2) SPD and confidence envelope*

## Unit 17 Exponential Regression

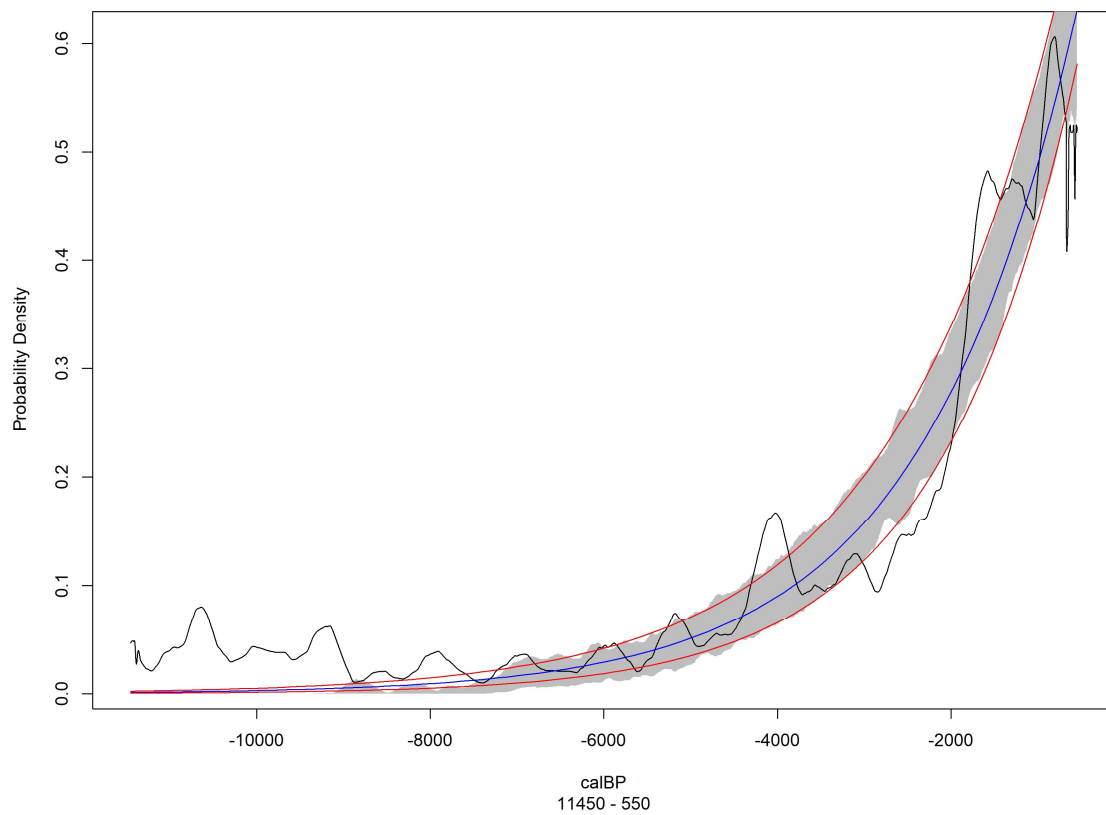
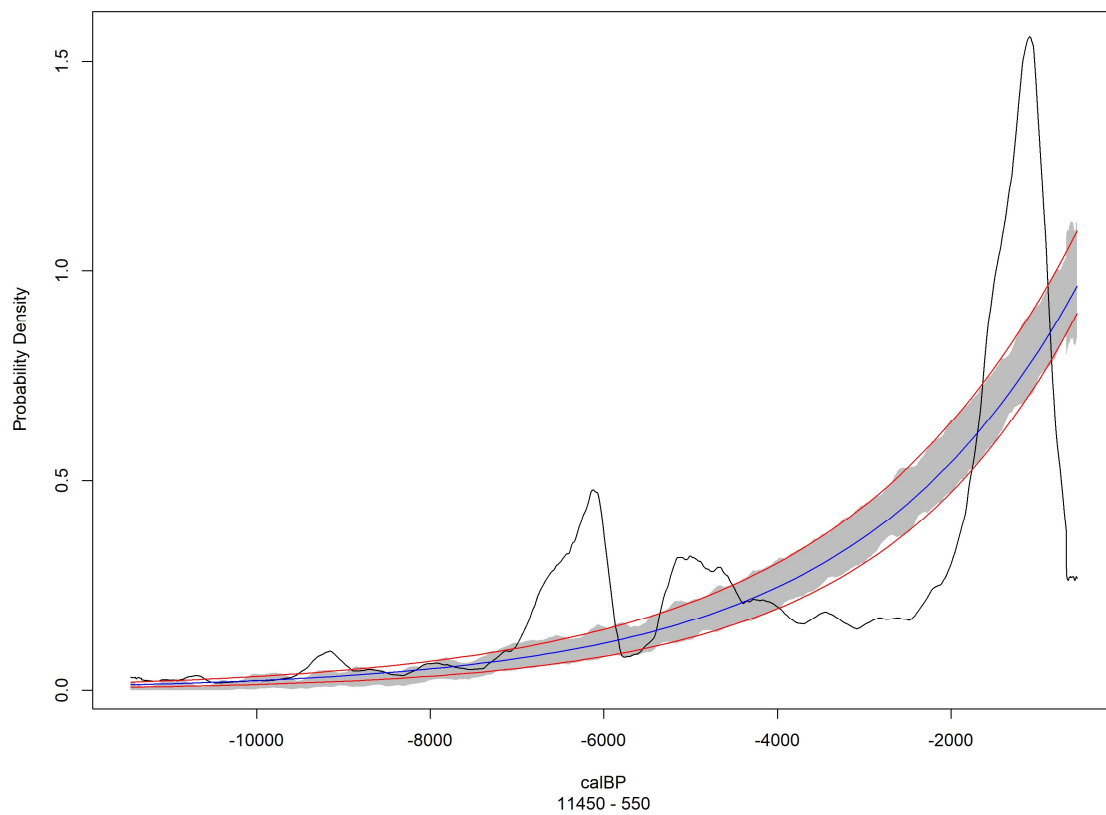


Figure AI.2 Exponential regressions for Nevada (Unit 17) SPD and confidence envelope

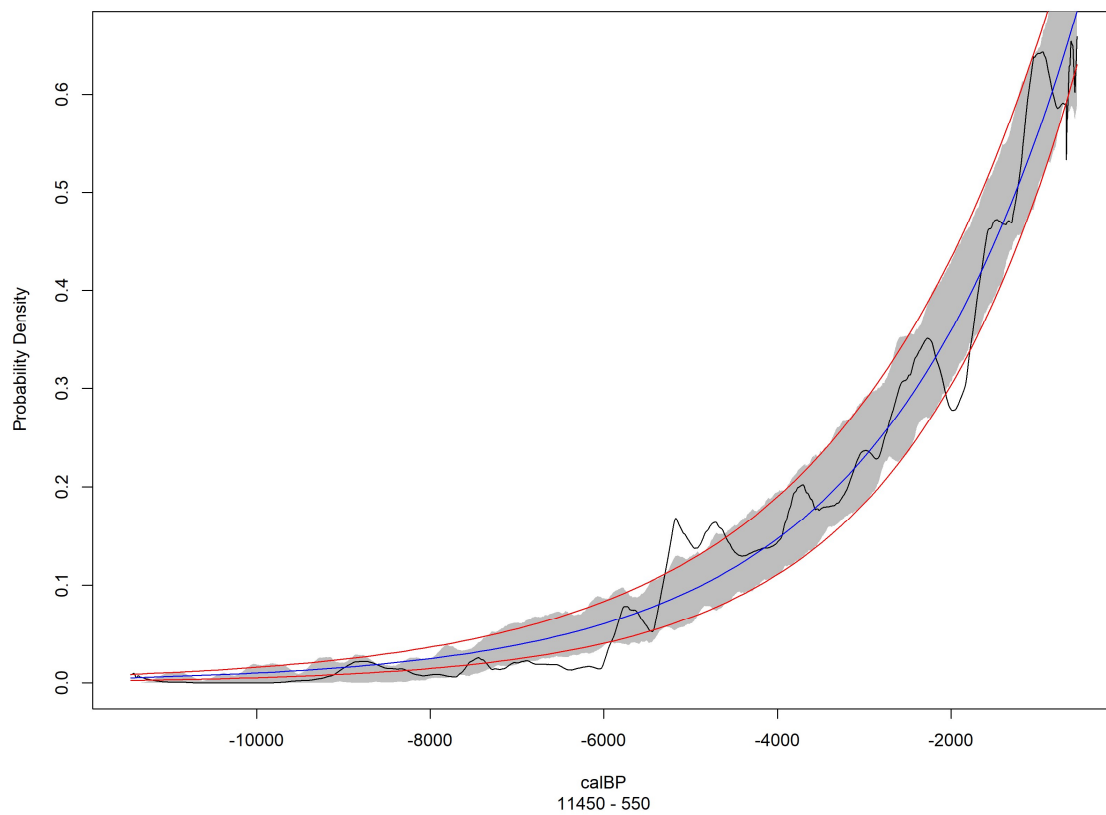


## Unit 30 Exponential Regression



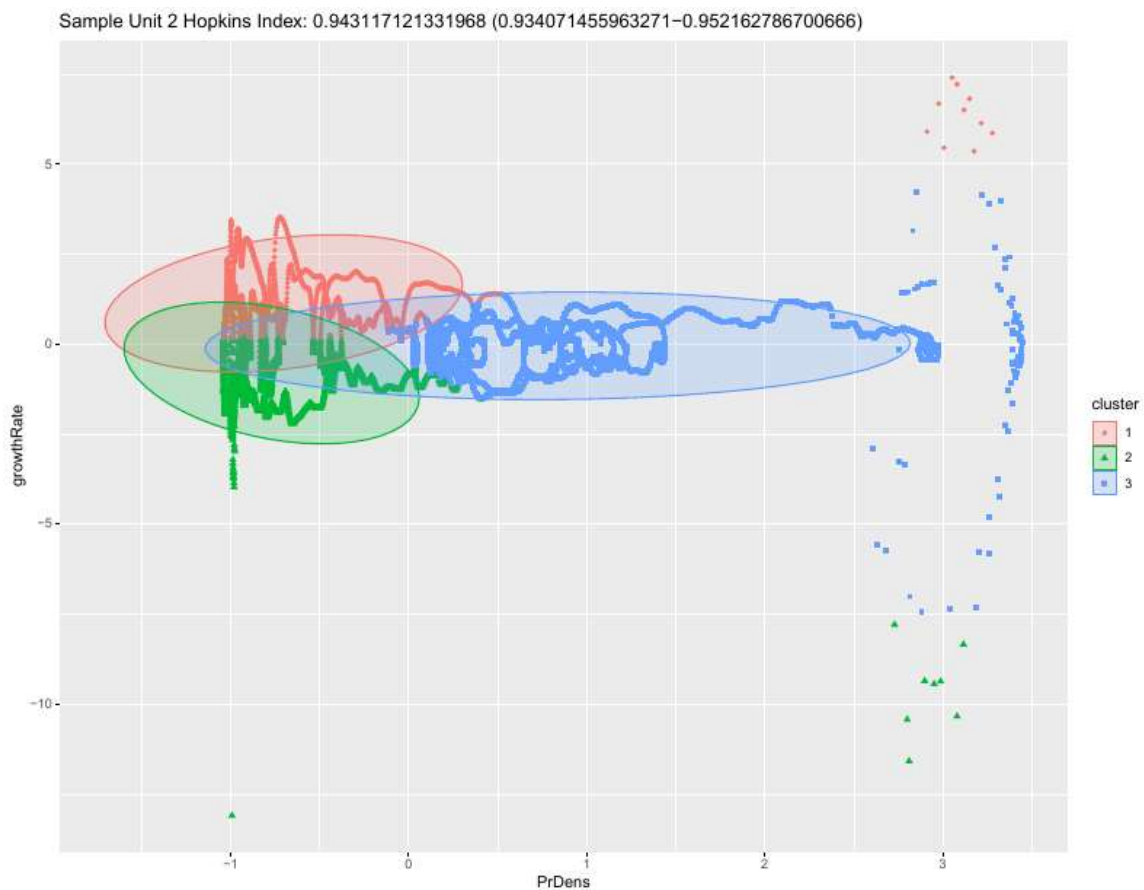
*Figure A1.3 Exponential regressions for Wyoming (Unit 30) SPD and confidence envelope*

## Unit 43 Exponential Regression



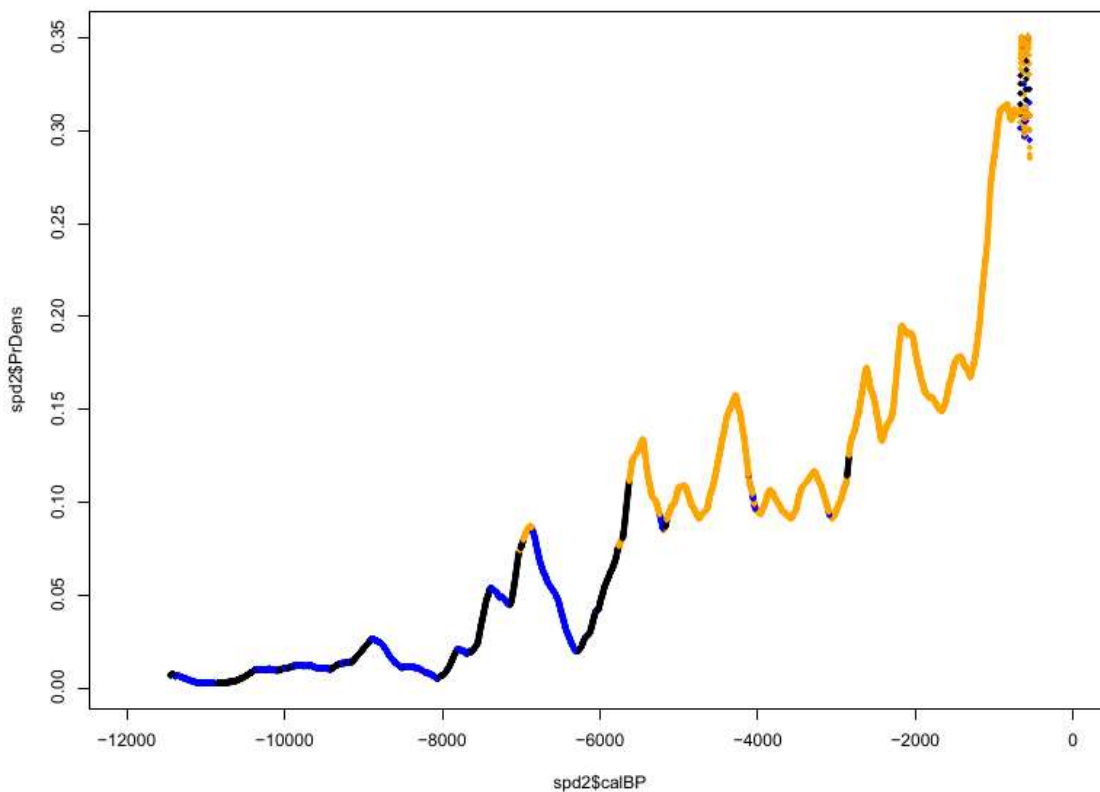
*Figure AI.4 Exponential regressions for PNW (Unit 43) SPD and confidence envelope*

## APPENDIX II: CLUSTER ANALYSES

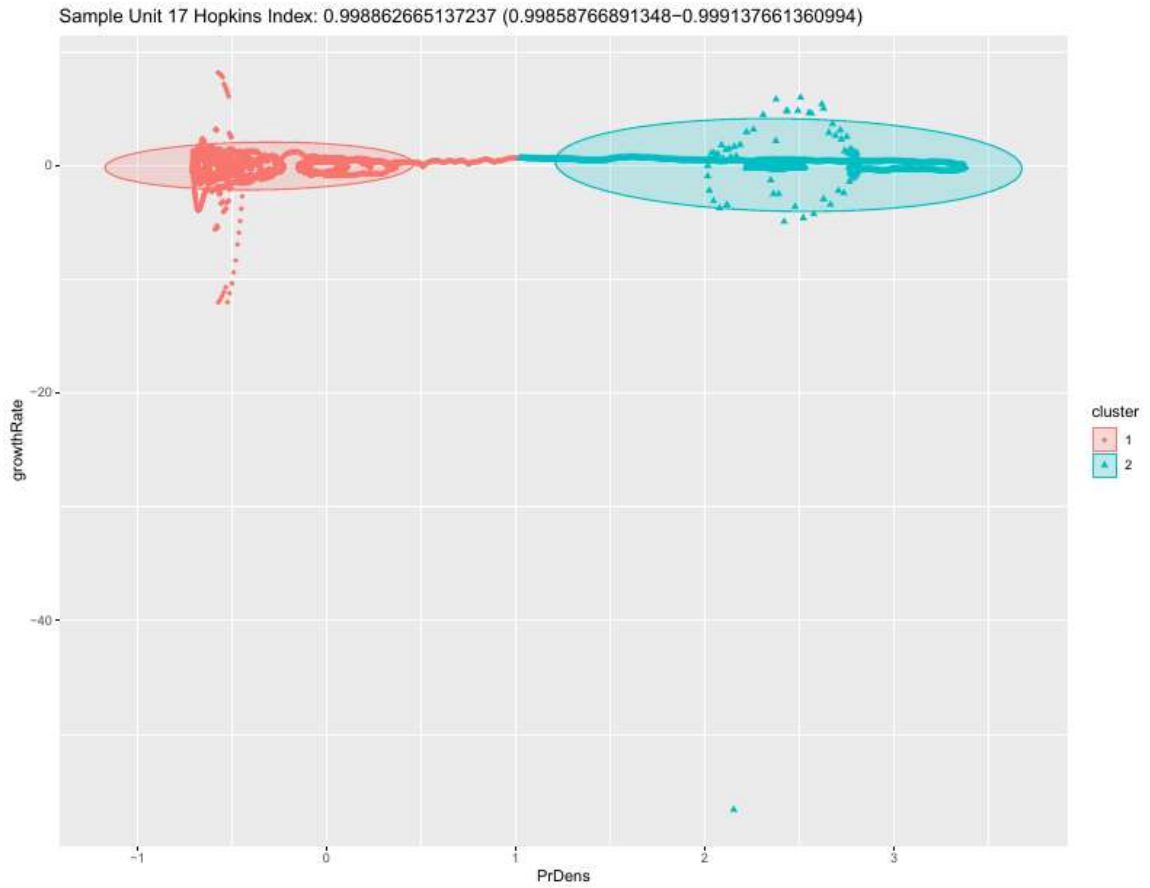


*Figure AII.1 Cluster analysis for Texas (Unit 2) Note cluster 2's (green) similarity in absolute magnitude to cluster 1 (red). The primary difference between cluster 1 and cluster 2 appears to be sign (positive or negative) associated with growth rate.*

Sample Unit 2 Hopkins Index: 0.943117121331968 (0.934071455963271–0.952162786700666)



*Figure AII.2 SPD for Texas (Unit 2) colored by cluster. Cluster 2 (blue) appears primarily to represent negative slopes within cluster 1 (black) region, so I treated this SPD as if only 2 clusters had been detected. Because clusters lack a clean delineation, selecting subset boundaries for this unit was somewhat subjective, so I picked a date representing the boundary for what appeared to be a large run (5630 calBP)*



*Figure AII.3 Cluster analysis for Nevada (Unit 17)—Clusters appear to be somewhat cleanly delineated*

Sample Unit 17 Hopkins Index: 0.998862665137237 (0.99858766891348–0.999137661360994)

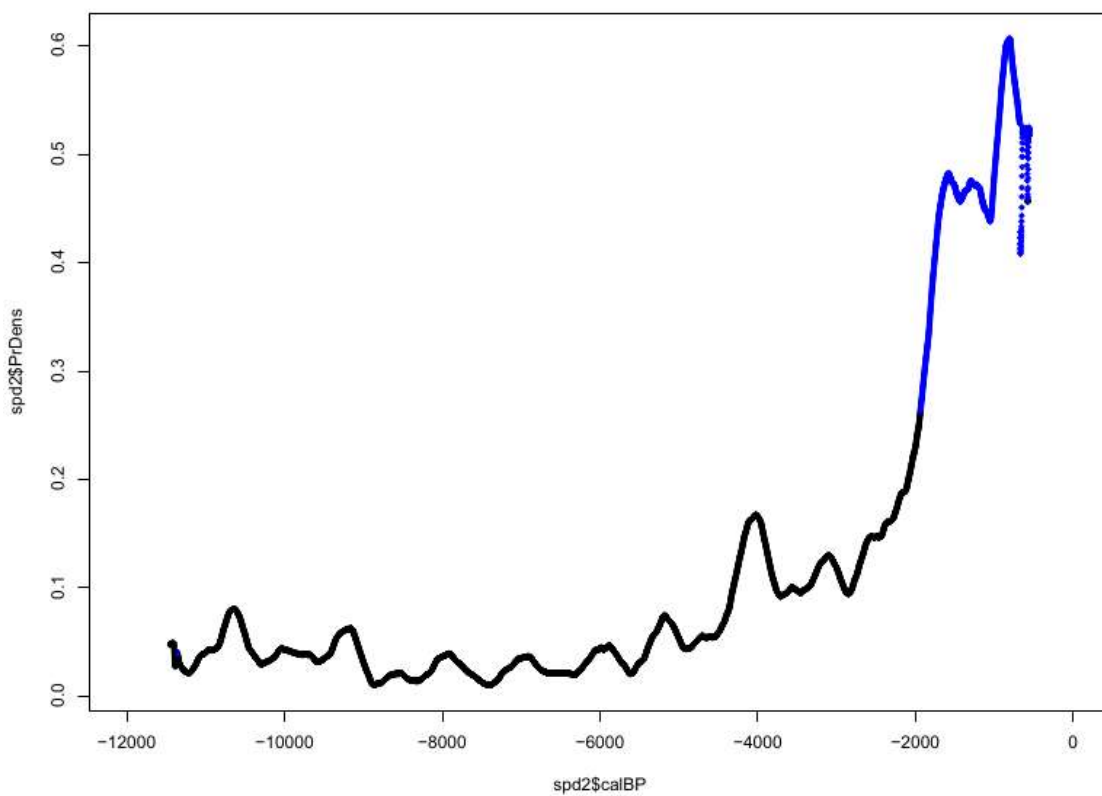
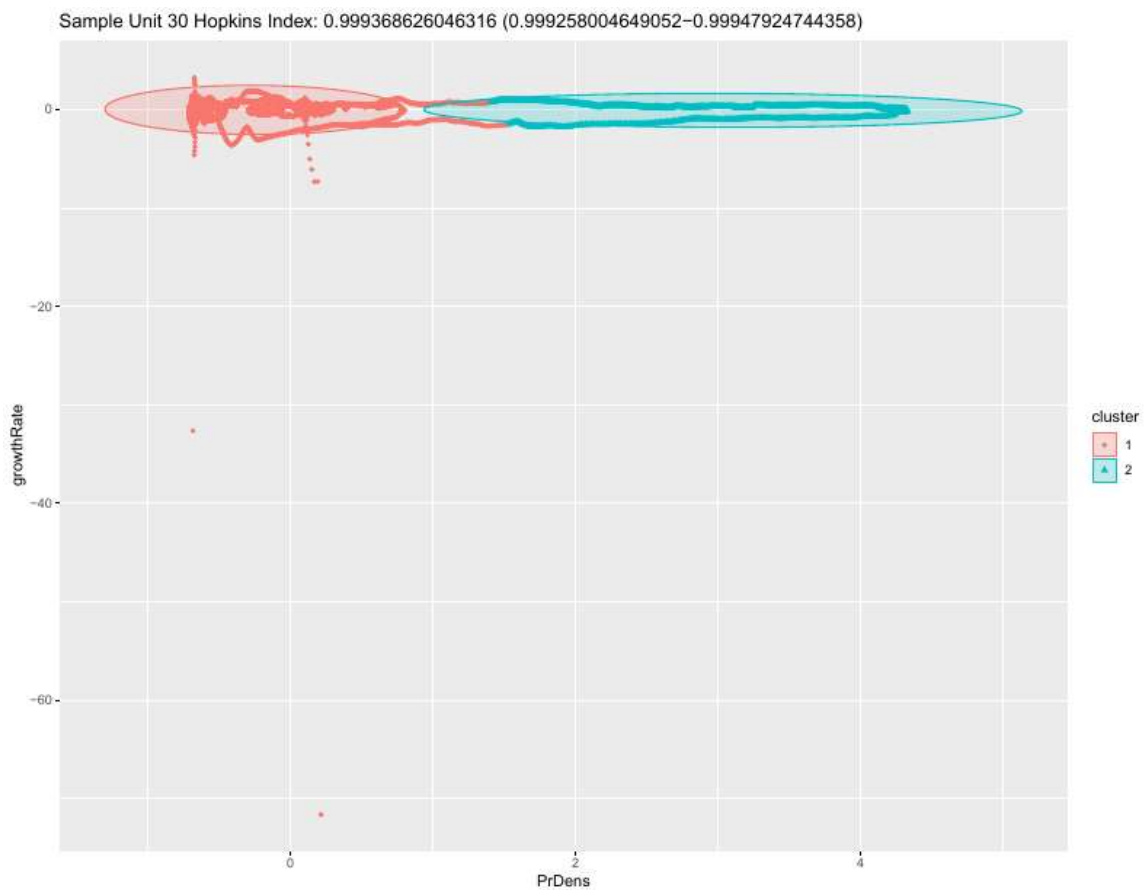


Figure AII.4 SPD for Nevada (Unit 17) colored by cluster



*Figure AII.5 Cluster analysis for Wyoming (Unit 30)—Clusters appear to be somewhat cleanly delineated*

Sample Unit 30 Hopkins Index: 0.999368626046316 (0.999258004649052-0.99947924744358)

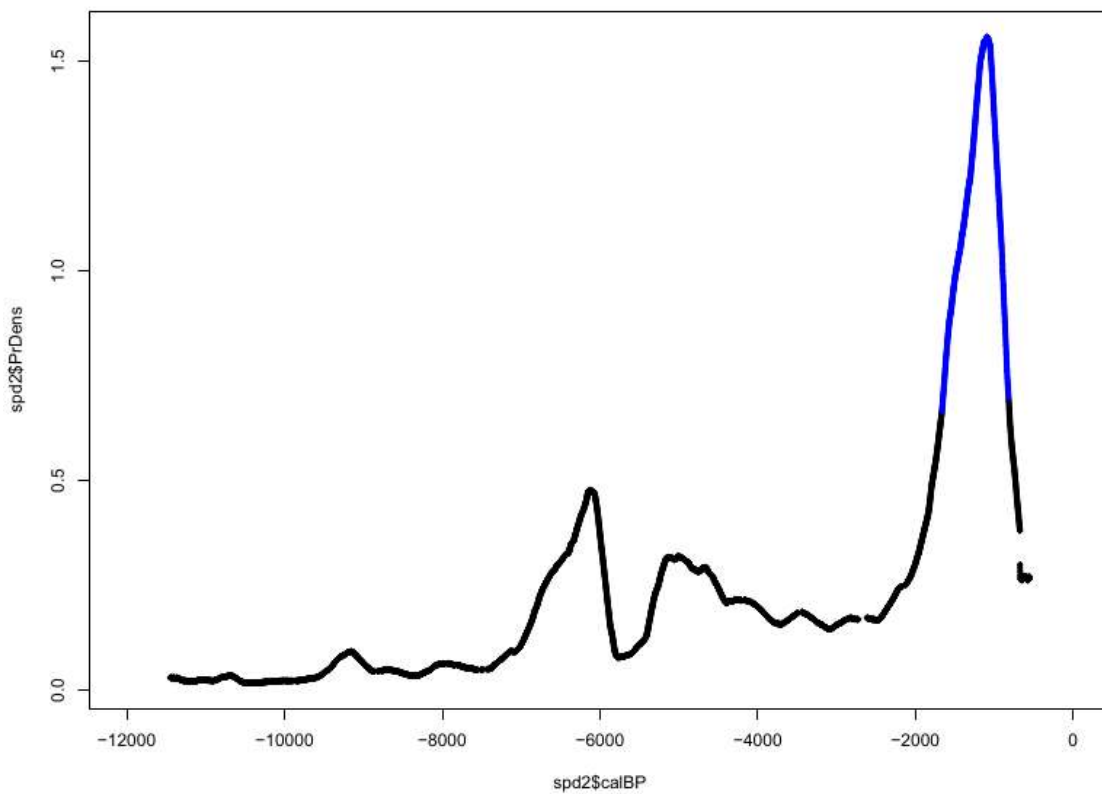


Figure AII.6 SPD for Wyoming (Unit 30) colored by cluster



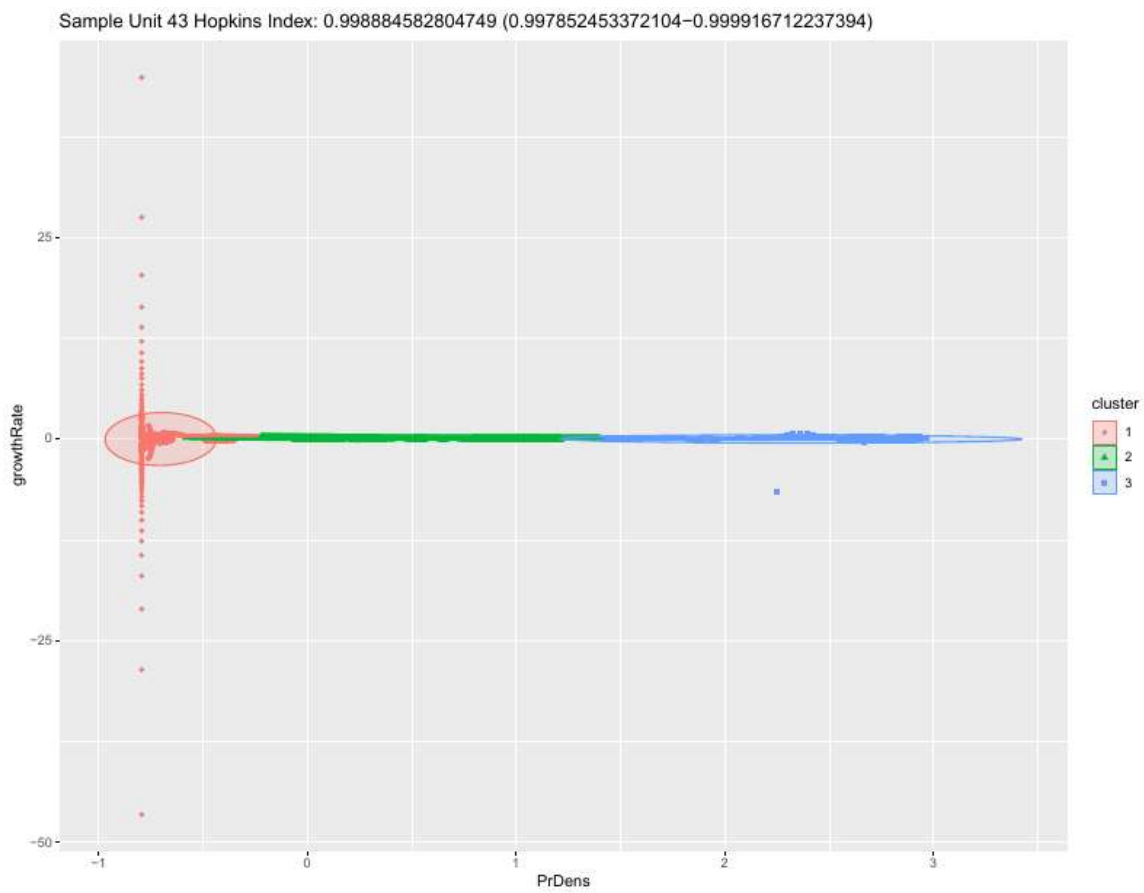


Figure AII.7 Cluster analysis for PNW (Unit 43)

Sample Unit 43 Hopkins Index: 0.998884582804749 (0.997852453372104-0.999916712237394)

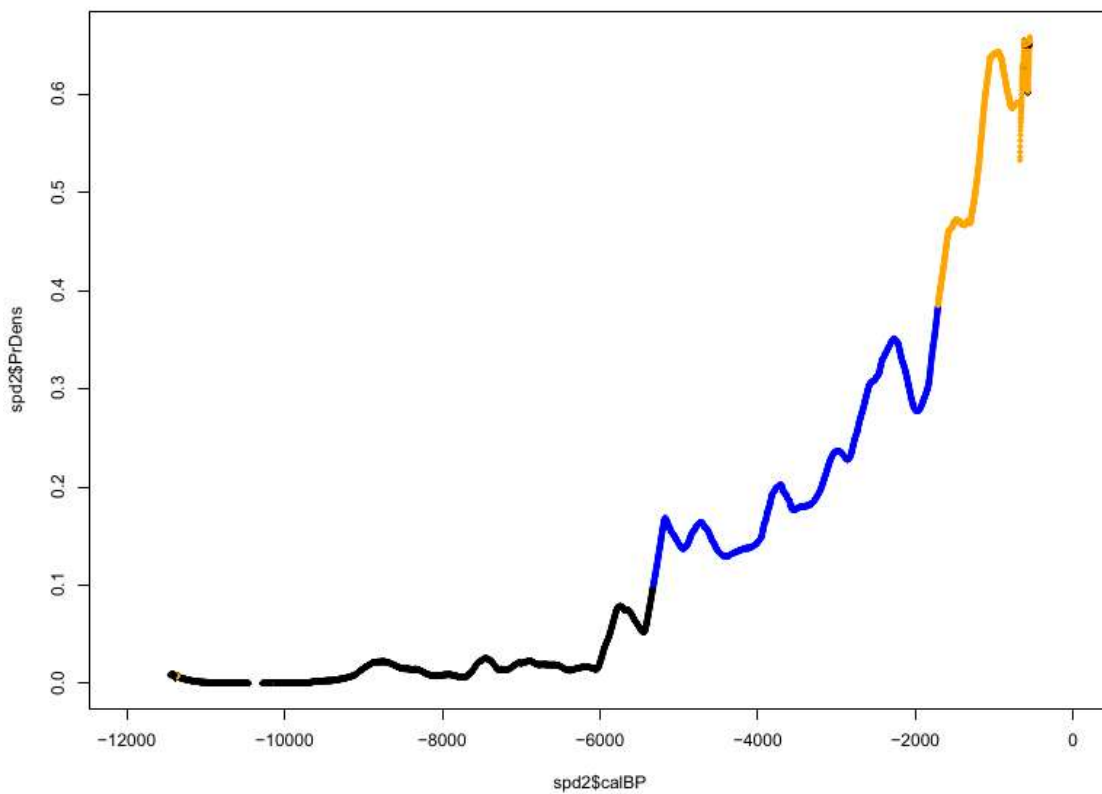
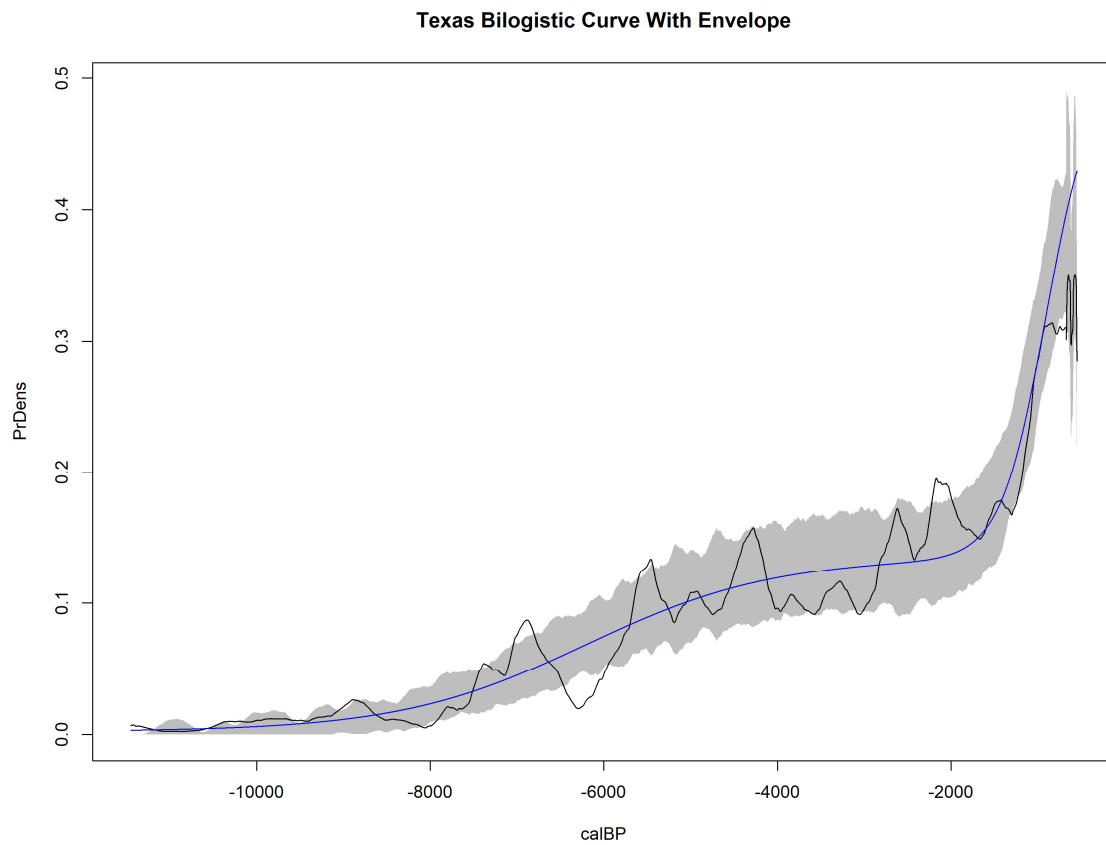
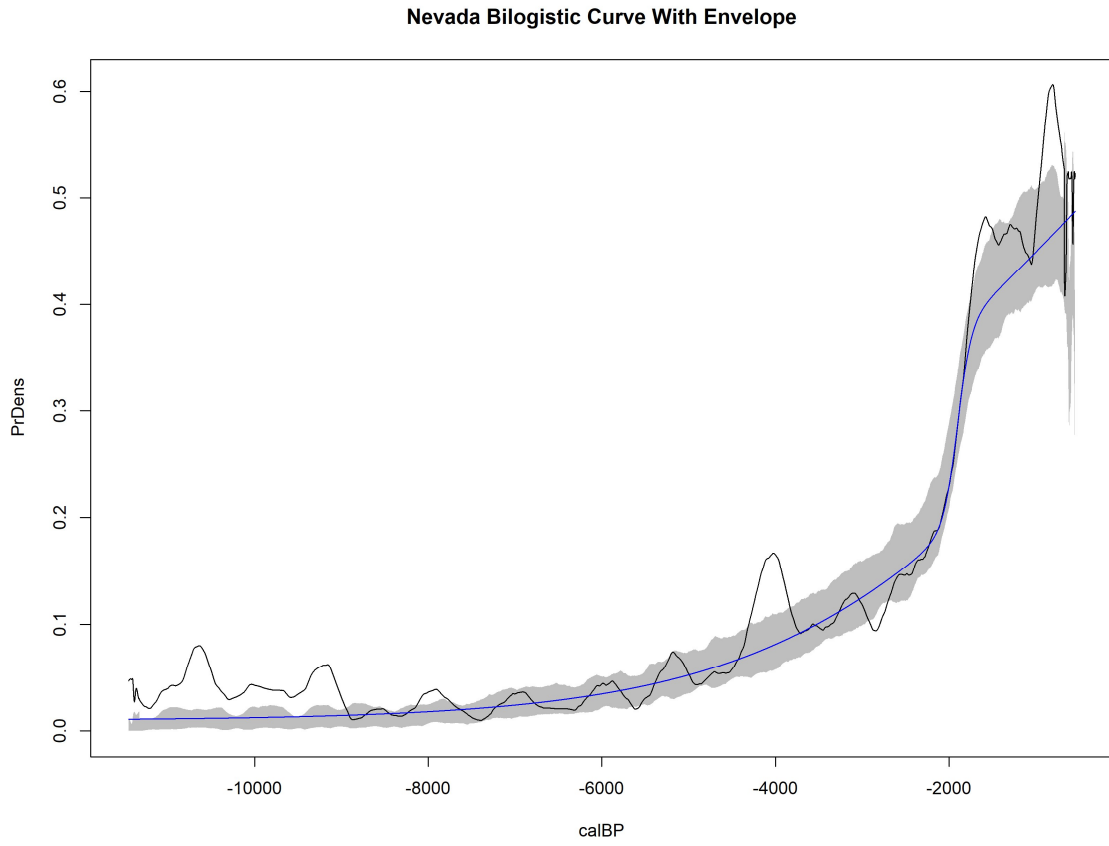


Figure AII.8 SPD for PNW (Unit 43) colored by cluster

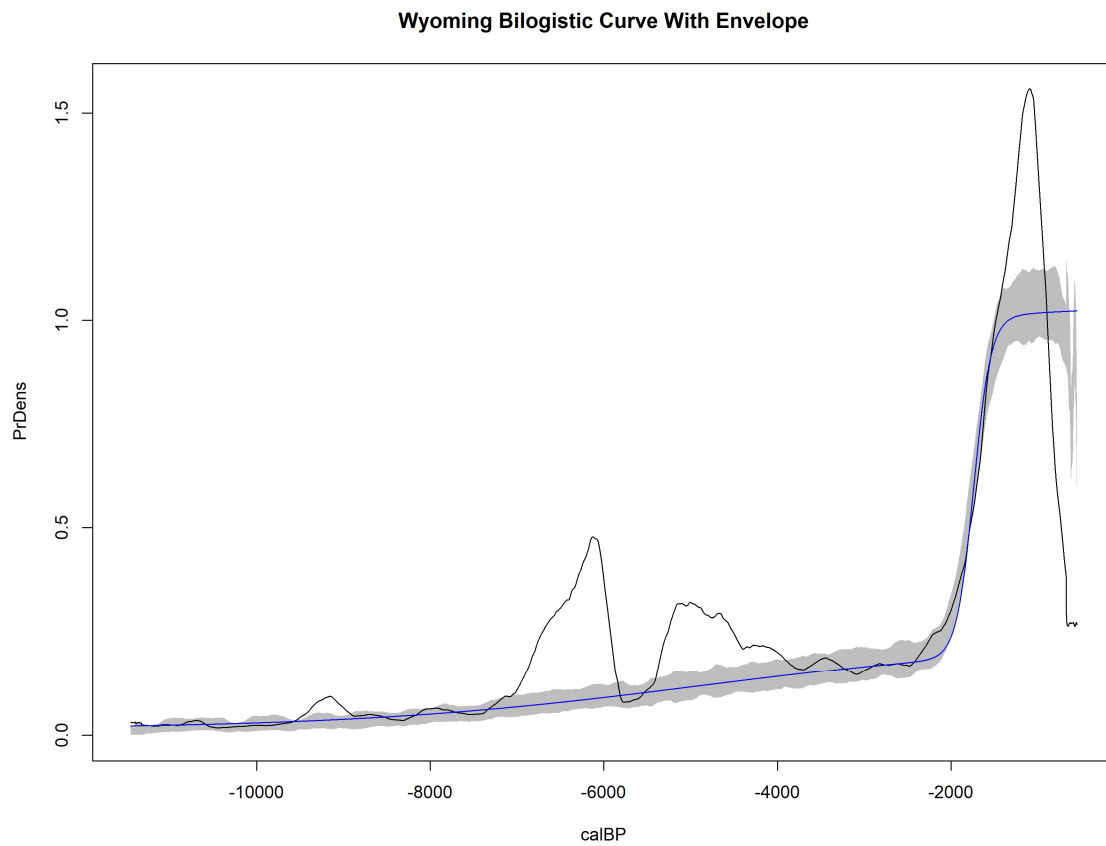
## APPENDIX III: MULTILOGISTIC CURVES WITH CONFIDENCE ENVELOPES



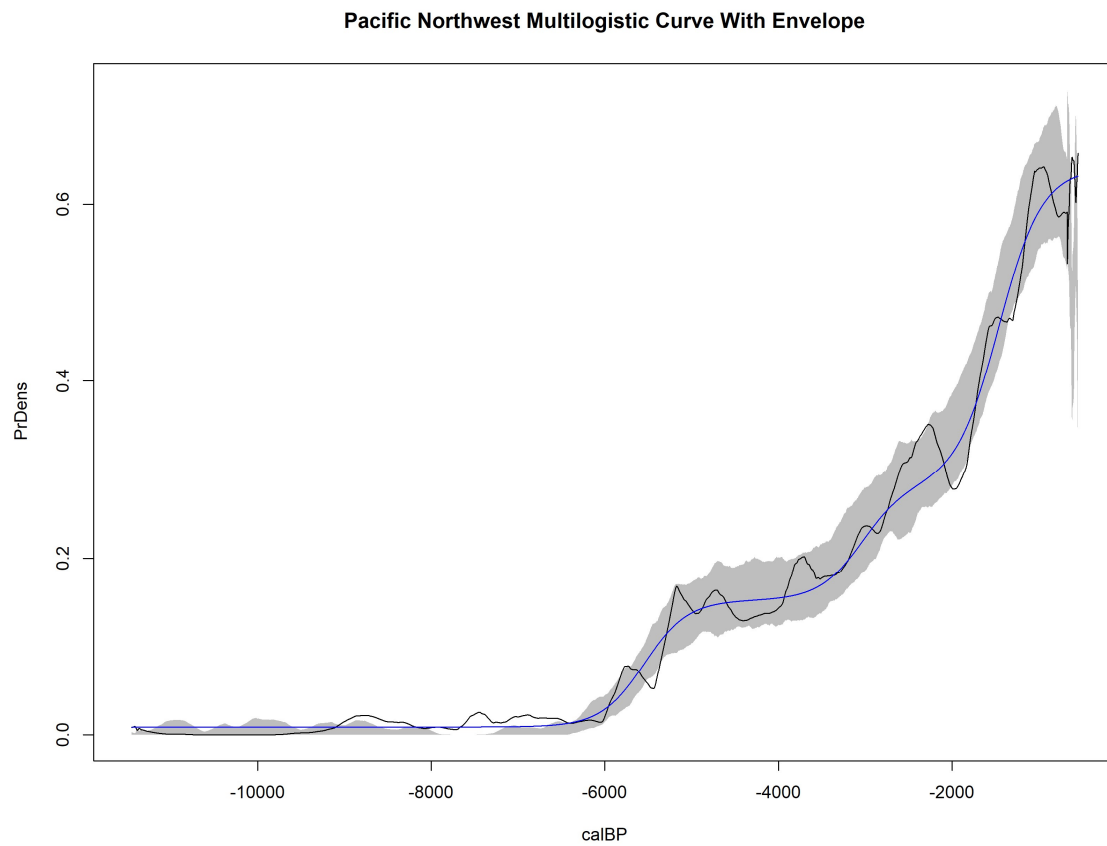
*Figure AIII.1 Bilogistic curve with confidence envelope fit to Texas (Unit 2)*



*Figure AIII.2 Bilogistic curve with confidence envelope fit to Nevada (Unit 17)*



*Figure AIII.3 Biogistic curve with confidence envelope fit to Wyoming (Unit 30)*



*Figure AIII.4 Multilogistic curve with confidence envelope fit to PNW (Unit 43)*

## APPENDIX IV: R CODE

```
library(foreach)
library(rcarbon)
library(drc)
currentPath <- "F:/"
SPDFolder <- paste(currentPath,
                   "SPD/",
                   sep = "")
pdfOutputs <- paste(currentPath,
                    "PDF/",
                    sep = "")
csvOutputs <- paste(currentPath,
                    "csvOutputs/",
                    sep = "")

if(!dir.exists(SPDFolder)){
  dir.create(SPDFolder)
}

if(!dir.exists(pdfOutputs)){
  dir.create(pdfOutputs)
}

if(!dir.exists(csvOutputs)){
  dir.create(csvOutputs)
}
```

```

#set a time range
timeRange <- c(11450,550)

boxLookup <- c(2,17,30,43)
boxLookup <- cbind(boxLookup, c("Texas", "Nevada", "Wyoming", "Pacific
Northwest"))
increment <- 3000

#get data from current folder
RawData <- read.csv(paste(currentPath,
                          "RawData(Scrubbed).csv",
                          sep = ""))

##### 2. Make Directory for all sample units with more than 200 lab numbers -----
Directory <- plyr::count(RawData,
                        vars = "Sbox")
names(Directory) <- c("Sbox",
                    "n")
Directory <- Directory[!(Directory$n < 200),]
Directory <- Directory[Directory$Sbox %in% c(2,
                                           17,
                                           30,
                                           43),]

#write.csv(Directory,paste(currentPath,"Directory_Sbox.csv",sep = ""),row.names =
FALSE)

##### 3. Calibrate each sampling unit recursively -----

```



#Sometimes r treats values within a dataframe in a way you cannot use. The lines ensure our calibration will work.

```
RawData$Age <- as.numeric(as.character(RawData$Age))
```

```
RawData <- RawData[!is.na(RawData$Age),]
```

```
RawData$Sd <- as.numeric(as.character(RawData$Sd))
```

```
RawData <- RawData[!(is.na(RawData$Sd)),]
```

```
RawData <- RawData[RawData$Sd > 0,]
```

```
RawData$LabID <- as.character(RawData$LabID)
```

##Turn each sampling unit into an element of a list.

```
SboxList <- vector("list", length(Directory$Sbox))
```

```
names(SboxList) <- Directory$Sbox
```

```
SboxList <- split(RawData[which(RawData$Sbox %in% Directory$Sbox),],
```

```
                  f = RawData[which(RawData$Sbox %in% Directory$Sbox),]$Sbox)
```

```
remove(RawData)
```

```
remove(Directory)
```

#For each sampling unit (sbox),

```
foreach(i = 1:length(SboxList)) %do% {
```

```
  Sbox <- data.frame(SboxList[[i]])
```

  #calibrate dates

```
  cptcall <- calibrate(x = Sbox$Age, errors = Sbox$Sd, normalised = F)
```

  #and generate SPDs using the exponential model

```
  SPD <- modelTest(cptcall,
```

```
                  errors = Sbox$Sd,
```

```

nsim = 1000,
bins = NA,
runm = 250,
timeRange = timeRange,
model = "exponential",
normalised = F,
raw = T,
edgeSize = -550,
ncores = 6,
verbose = T,
predgrid = predGrid)$result

#Exploring things like slope is easier with negative calBP values
SPD$calBP <- -abs(SPD$calBP)
SPD$$Sbox <- data.frame(SboxList[[i]])$Sbox[1]

#Save the SPD
write.csv(SPD, file = paste(SPDFolder, "Box_", SPD$$Sbox[1],
"_SPD_and_envelope_(edgesize550).csv", sep = ""), row.names = F)

#Plot each SPD at each given value to its own file
pdf(file = paste(pdfOutputs,
SPD$$Sbox[1],
"_exp_plot(edgesize550).pdf",
sep = ""),
height = 8.5,
width = 11,
title = paste("SPD Sample Unit ",
SPD$$Sbox[1],

```

```

        sep = ""))
plot(SPD$PrDens~SPD$calBP,
     type = "lines",
     main = paste("SPD Sample Unit ",
                  SPD$Sbox[1],
                  sep = ""),
     xlab = "CalBP",
     ylab = "Probability Density")
polygon(c(SPD$calBP, rev(SPD$calBP)),
        c(SPD$lo, rev(SPD$hi)),
        col = "grey",
        lty = 0)
lines(SPD$PrDens~SPD$calBP,
      type = "lines")
abline(h = flat, col = "blue")
dev.off()

#Print a line for each completed samping unit (optional)
cat(paste("\nExtraction for Sampling Unit ",
          data.frame(SboxList[[i]])$Sbox[1],
          " (",
          i,
          "/",
          length(unique(SboxList)),
          ")"),
    "complete\n\n"))
}

```

```

#get slope data. I originally wrote this code intending to explore timescale
#sensitivity of exponential slopes; I have left the code

SPDList <- list.files(path = SPDFolder,
                     pattern = ".*[0-9]_SPD_and_envelope_.edge.*.csv")

slopeDataFileNames <- gsub("_and_envelope", "_slope_data", SPDList)
pdfFileNames <- gsub("_and_envelope", "nlsRegression", SPDList)
pdfFileNames <- gsub(".csv", ".pdf", pdfFileNames)

#assign files a complete location
SPDList <- paste(SPDFolder, SPDList, sep = "")
slopeDataFileNames <- paste(csvOutputs, slopeDataFileNames, sep = "")
pdfFileNames <- paste(pdfOutputs, pdfFileNames, sep = "")

for(i in 1:length(SPDList)) {
  if(i == 1) {rm(list = c("slopeData", "SD1"))}

  #load the spd
  SPD <- read.csv(SPDList[i])

  #working in years relative to present (1950) is easier to interpret than
  #absolute years before present; specifically, we want positive slopes to
  #represent growth
  SPD$calBP <- -abs(SPD$calBP)

  #change the late value of the time range to the time of maximum population

```

```
# timeRange <- c(-12000, min(-250,SPD$calBP[SPD$PrDens == max(SPD$PrDens)]))
# SPD <- SPD[SPD$calBP < timeRange[2],]
```

```
j <- 1
```

```
while(j < length(SPD[,1])){
```

```
  cat(paste("\n", j))
```

```
  #calBP includes numbers that are too high for exp() to handle, so we scale
  #it down. Note that the resulting slope must be rescaled.
```

```
  currentNLS <- nls(PrDens ~ exp(a + (calBP / 1000) * b),
    data = SPD[(j:length(SPD$calBP)),],
    start = list(a = 0.5, b = 0.5),
    nls.control(maxiter = 2048))
```

```
  NLSAicc <- AICcmodavg::AICc(currentNLS)
```

```
  NLSBIC <- BIC(currentNLS)
```

```
  currentNLSHigh <- nls(hi ~ exp(a + (calBP / 1000) * b),
    data = SPD[(j:length(SPD$calBP)),],
    start = list(a = 0.5, b = 0.5),
    nls.control(maxiter = 2048))
```

```
  currentNLSLow <- nls(lo ~ exp(a + (calBP / 1000) * b),
    data = SPD[(j:length(SPD$calBP)),],
    start = list(a = 0.5, b = 0.5),
    nls.control(maxiter = 2048))
```

```
  SD1 <- cbind(sampleUnit = SPD$$Sbox[i],
```

```
    unitName = boxLookup[boxLookup[,1] %in% SPD$$Sbox[i],2],
```

```
    timeBegin = -SPD[j,1],
```

```

timeEnd = -timeRange[2],
nlsTransl = (currentNLS$m$getPars()[1]),
nlsLowTransl = (currentNLSLow$m$getPars()[1]),
nlsHighTransl = (currentNLSHigh$m$getPars()[1]),
nlsTransf = (currentNLS$m$getPars()[2]),
nlsLowTransf = (currentNLSLow$m$getPars()[2]),
nlsHighTransf = (currentNLSHigh$m$getPars()[2]),
AICc = NLSAicc,
BIC = NLSBIC)
ifelse(!exists("slopeData"),
      slopeData <- SD1,
      slopeData <- rbind(slopeData,
                          SD1))
slopeData <- as.data.frame(slopeData)
j <- j+increment
}

if(!exists("slopeData1")) {

  slopeData1 <- slopeData

} else {

  slopeData1 <- rbind(slopeData1, slopeData)

}

# return(as.list(slopeData))
write.csv(slopeData, slopeDataFileNames[i], row.names = F)

```

```

pdf(file = pdfFileNames[i],height = 8.5, width = 11)

plot(SPD$PrDens~SPD$calBP,
     type = "lines",
     xlab = "calBP",
     ylab = "Probability Density",
     main = paste("Unit ",
                  slopeData$sampleUnit[1],
                  # " h",
                  # slopeData$h[1],
                  " Exponential Regression",
                  sep = ""),
     sub = paste(max(abs(SPD$calBP)),
                 " - ",
                 min(abs(SPD$calBP)),
                 sep = ""))

polygon(c(SPD$calBP, rev(SPD$calBP)),
        c(SPD$lo, rev(SPD$hi)),
        col = "grey",
        lty = 0)

lines(SPD$PrDens~SPD$calBP,
      type = "lines")

curve(exp(as.numeric(slopeData[1,5])+as.numeric(slopeData[1,8])*x/1000), col =
"blue", add = T, from = min(SPD$calBP), to = max(SPD$calBP))

curve(exp(as.numeric(slopeData[1,6])+as.numeric(slopeData[1,9])*x/1000), col =
"red", add = T, from = min(SPD$calBP), to = max(SPD$calBP))

curve(exp(as.numeric(slopeData[1,7])+as.numeric(slopeData[1,10])*x/1000), col =
"red", add = T, from = min(SPD$calBP), to = max(SPD$calBP))

```

```

dev.off()

png(file = imageFileNames[i],height = 8.5, width = 11, units = "in", res = 300)

plot(SPD$PrDens~SPD$calBP,
     type = "lines",
     xlab = "calBP",
     ylab = "Probability Density",
     main = paste("Unit ",
                  slopeData$sampleUnit[1],
                  # " h",
                  # slopeData$h[1],
                  " Exponential Regression",
                  sep = ""),
     sub = paste(max(abs(SPD$calBP)),
                 " - ",
                 min(abs(SPD$calBP)),
                 sep = ""))

polygon(c(SPD$calBP, rev(SPD$calBP)),
        c(SPD$lo, rev(SPD$hi)),
        col = "grey",
        lty = 0)

lines(SPD$PrDens~SPD$calBP,
      type = "lines")

curve(exp(as.numeric(slopeData[1,5])+as.numeric(slopeData[1,8])*x/1000), col =
"blue", add = T, from = min(SPD$calBP), to = max(SPD$calBP))

curve(exp(as.numeric(slopeData[1,6])+as.numeric(slopeData[1,9])*x/1000), col =
"red", add = T, from = min(SPD$calBP), to = max(SPD$calBP))

curve(exp(as.numeric(slopeData[1,7])+as.numeric(slopeData[1,10])*x/1000), col =
"red", add = T, from = min(SPD$calBP), to = max(SPD$calBP))

dev.off()

```



```

rm("slopeData")

}

BICexp <- slopeData1[which(slopeData1$timeBegin == 11450),c(1,ncol(slopeData1))]
write.csv(BICexp, paste(csvOutputs, "BICexp.csv", sep = ""), row.names = F)

fileNumbers <- gsub("\\D", "", SPDLlist)
fileNumbers <- gsub("550", "", fileNumbers)

buildModel <- function(clusterCount, x_axis, y_axis, starter=NA) {

  formulae <- c(paste("limit", 1:clusterCount, "/(1+exp(-1*slope", 1:clusterCount,
"*(xmid_", 1:clusterCount, "-", x_axis, "))))", sep = ""))
  formulae <- gsub("limit1", "(limit1 - d)", formulae)
  formulae <- append(formulae, "d")

  if(!missing(starter)) {

    for(i in 1:length(starter[,1])){
      formulae <- gsub(starter[i,1], starter[i,2], formulae)
    }

  }

  model <<- as.formula(paste(y_axis, "~", paste(formulae, collapse = "+")))

}

```

```

clusterCounts <- c(2,2,2,3)
robust <- c("mean", "median", "lms", "lts")
modelList <- list()
predGrid <- list()
paramList <- list()
robustList <- matrix(nrow = 4, ncol = 3)
BICdf <- matrix(nrow = 4, ncol = 2)
# i <- 2
for(i in 1:4) {

  robust1 <- expand.grid(1:4, 1:4)

  if(i == 4) {

    robust1 <- expand.grid(1:4, 1:4, 1:4)

  }

  #load SPD
  SPD <- read.csv(clusterFiles[i])
  clusterFiles[i]
  firstXmid <- c(-1937, -5630, -1664, -5321)[i]
  secondXmid <- -1713

  #Interpreting regression values is simpler if the regression is run on negative
  #calBP values. I've used the negative absolute value just in case I forget and
  #run this line twice; this way I don't inadvertently flip the sign back to
  #positive.
  for(j in 1:nrow(robust1)){

```

```

if(j == 1) {

  methodSelection <- numeric(length = nrow(robust1))

}

SPD$scalBP <- -abs(SPD$scalBP)
SPD1 <- SPD[1:which(SPD$scalBP == (firstXmid+1000)),]
# plot(SPD1$PrDens~SPD1$scalBP, type = "lines")

drm1 <- drm(PrDens~calBP, fct = L.4(names = c("slope1", "d", "limit1", "xmid_1")),
data = SPD1, robust = robust[robust1[j,1]], lowerl = c(-Inf, 0, 0, -Inf))

drmPred1 <- data.frame(calBP = SPD$scalBP)

coeffCount <- 1:length(names(drm1$coefficients))

eval(parse(text =
paste("drmPred1$\\",gsub(":.*", "", names(drm1$coefficients[coeffCount])), "\\ " <-
(drm1$coefficients["", coeffCount, "]), sep = "")))

drmPred1$PrDens <- predict(drm1, drmPred1)

SPD1 <- SPD[which(SPD$scalBP == (firstXmid-1000)):ifelse(clusterCounts[i] > 2,
which(SPD$scalBP == (secondXmid + 1000)), length(SPD$scalBP)),]

drm1Effects <- drmPred1[which(drmPred$scalBP %in% SPD1$scalBP),]

SPD1$PrDens <- (SPD1$PrDens - drm1Effects$PrDens)

drm2 <- drm(PrDens~calBP, fct = L.3(names = c("slope2", "limit2", "xmid_2")), data
= SPD1, robust = robust[robust1[j,2]], lowerl = c(-Inf, 0, -Inf))

drmPred2 <- data.frame(calBP = SPD$scalBP)

coeffCount <- 1:length(names(drm2$coefficients))

eval(parse(text =
paste("drmPred2$\\",gsub(":.*", "", names(drm2$coefficients[coeffCount])), "\\ " <-
(drm2$coefficients["", coeffCount, "]), sep = "")))

```

```

drmPred2$PrDens <- predict(drm2, drmPred2)

if(clusterCounts[i] > 2){

  SPD1 <- SPD[which(SPD$calBP == (secondXmid-1000)):length(SPD$calBP),]
  drm1Effects <- drmPred1[which(drmPred1$calBP %in% SPD1$calBP),]
  drm2Effects <- drmPred2[which(drmPred2$calBP %in% SPD1$calBP),]
  SPD1$PrDens <- (SPD1$PrDens - (drm1Effects$PrDens + drm2Effects$PrDens))

  drm3 <- drm(PrDens~calBP, fct = L.3(names = c("slope3", "limit3", "xmid_3")), data
= SPD1, robust = robust[robust1[j,3]], lower1 = c(-Inf, 0, -Inf))
  drmPred3 <- data.frame(calBP = SPD$calBP)
  drmPred3 <- data.frame(calBP = SPD$calBP)
  coeffCount <- 1:length(names(drm3$coefficients))

  eval(parse(text =
paste("drmPred3$\"",gsub(":.*", "", names(drm3$coefficients[coeffCount])), "\" <-
(drm3$coefficients[", coeffCount, "]", sep = "")))
  drmPred3$PrDens <- predict(drm3, drmPred3)

}

#We want the AICc for the summed drc objects. To do this, I build the summed
#model and run an NLS, forcing the lower asymptote to the value regressed with
#drm

FromTo <- 1:clusterCounts[i]

# startBuilder <- c("starter <- c(",
#                   paste("limit", FromTo, " = unname(drm", FromTo,

```

```

"$coefficients[which(names(drm", FromTo, "$coefficients)==\"d:(Intercept)\")],", sep =
""),
#       paste("xmid_", FromTo, " = unname(drm", FromTo,
"$coefficients[which(names(drm", FromTo, "$coefficients)==\"e:(Intercept)\")],", sep =
""),
#       paste("slope", FromTo, " = unname(drm", FromTo,
"$coefficients[which(names(drm", FromTo, "$coefficients)==\"b:(Intercept)\")],", sep =
""),
#       ")")
startBuilder <- c("starter <- c(",
                 paste("drm", FromTo, "$coefficients," , sep = ""),
                 ")")
startBuilder[(length(startBuilder)-1)] <- gsub("$", "",
startBuilder[(length(startBuilder)-1)])
eval(parse(text = startBuilder))
names(starter) <- gsub(":.*", "", names(starter))
starter <- starter[-which(names(starter) == "d")]
starter1 <- data.frame(names = names(starter), value = unname(starter))

model <- buildModel(clusterCounts[i], "calBP", "PrDens", starter = starter1)
# model

nlsFinal <- nls(model, data = SPD, start = list(d = unname(drm1$coefficients[2])),
upper = list(d = unname(drm1$coefficients[2])), lower = list(d =
unname(drm1$coefficients[2])), algorithm = "port")
plot(SPD$PrDens~SPD$calBP, type = "lines") +
  lines(nlsFinal$m$fitted()~SPD$calBP, col = "blue")
# plot(nlsFinal$m$fitted()~SPD$calBP, col = "blue", type = "lines")

cat(paste("\n\n", robust1[j,1], robust1[j,2], ifelse(ncol(robust1) == 3, robust1[j,3], ""), "
BIC: ",(BIC(nlsFinal)+(log(nrow(SPD))*3*clusterCounts[i])), sep = ""))

```

```

methodSelection[j] <- (BIC(nlsFinal) + (log(nrow(SPD))*3*clusterCounts[i]))

if(j == nrow(robust1)){

  SPD1 <- SPD[1:which(SPD$calBP == (firstXmid+1000)),]

  plot(SPD1$PrDens~SPD1$calBP, xlim = c(-12000,0), ylim = c(0,
max(SPD$PrDens)), type = "lines", col = "black")

  # plot(SPD1$PrDens~SPD1$calBP, type = "lines")

  # drm1 <- drm(PrDens~calBP, fct = L.4(), data = SPD1, robust =
robust[robust1[which(methodSelection == min(methodSelection)),1]])

  drm1 <- drm(PrDens~calBP, fct = L.4(names = c("slope1", "d", "limit1", "xmid_1")),
data = SPD1, robust = robust[robust1[which(methodSelection ==
min(methodSelection)),1]], lowerl = c(-Inf, 0, 0, -Inf))

  drmPred1 <- data.frame(calBP = SPD$calBP)

  # drmPred1$b <- drm1$coefficients[1]

  # drmPred1$c <- drm1$coefficients[2]

  # drmPred1$d <- drm1$coefficients[3]

  # drmPred1$e <- drm1$coefficients[4]

  coeffCount <- 1:length(names(drm1$coefficients))

  eval(parse(text =
paste("drmPred1$",gsub(":".*, "", names(drm1$coefficients[coeffCount])), "\\" <-
(drm1$coefficients["", coeffCount, "]), sep = "")))

  drmPred1$PrDens <- predict(drm1, drmPred1)

  SPD1 <- SPD[which(SPD$calBP == (firstXmid-1000)):ifelse(clusterCounts[i] > 2,
which(SPD$calBP == (secondXmid + 1000)), length(SPD$calBP)),]

  drm1Effects <- drmPred1[which(drmPred1$calBP %in% SPD1$calBP),]

  SPD1$PrDens <- (SPD1$PrDens - drm1Effects$PrDens)

  lines(SPD1$PrDens~SPD1$calBP, xlim = c(-12000,0), ylim = c(0,

```

```

max(SPD$PrDens)), type = "lines", col = "blue")

# drm2 <- drm(PrDens~calBP, fct = L.3(), data = SPD1, robust =
robust[robust1[which(methodSelection == min(methodSelection)),2]])

drm2 <- drm(PrDens~calBP, fct = L.3(names = c("slope2", "limit2", "xmid_2")), data
= SPD1, robust = robust[robust1[which(methodSelection == min(methodSelection)),2]],
lowerl = c(-Inf, 0, -Inf))

drmPred2 <- data.frame(calBP = SPD$calBP)

coeffCount <- 1:length(names(drm2$coefficients))

eval(parse(text =
paste("drmPred2$\"",gsub(":.*", "", names(drm2$coefficients[coeffCount])), "\" <-
(drm2$coefficients[", coeffCount, "]", sep = "")))

# drmPred2$b <- drm2$coefficients[1]
# drmPred2$d <- drm2$coefficients[2]
# drmPred2$e <- drm2$coefficients[3]
drmPred2$PrDens <- predict(drm2, drmPred2)

if(clusterCounts[i] > 2){

SPD1 <- SPD[which(SPD$calBP == (secondXmid-1000)):length(SPD$calBP),]
drm1Effects <- drmPred1[which(drmPred1$calBP %in% SPD1$calBP),]
drm2Effects <- drmPred2[which(drmPred2$calBP %in% SPD1$calBP),]
SPD1$PrDens <- (SPD1$PrDens - (drm1Effects$PrDens + drm2Effects$PrDens))

lines(SPD1$PrDens~SPD1$calBP, xlim = c(-12000,0), ylim = c(0,
max(SPD$PrDens)), type = "lines", col = "orange")

drm3 <- drm(PrDens~calBP, fct = L.3(names = c("slope3", "limit3", "xmid_3")),
data = SPD1, robust = robust[robust1[which(methodSelection ==
min(methodSelection)),3]], lowerl = c(-Inf, 0, -Inf))

drmPred3 <- data.frame(calBP = SPD$calBP)

```

```

coeffCount <- 1:length(names(drm3$coefficients))

eval(parse(text =
paste("drmPred3$\\" ,gsub(":.*", "", names(drm3$coefficients[coeffCount])), "\\" <-
(drm3$coefficients[" , coeffCount, "]", sep = "")))

# drmPred3$b <- drm3$coefficients[1]
# drmPred3$d <- drm3$coefficients[2]
# drmPred3$e <- drm3$coefficients[3]

drmPred3$PrDens <- predict(drm3, drmPred3)

# plot(SPD1$PrDens~SPD1$calBP, xlim = c(-12000,0), ylim = c(0,
max(SPD$PrDens)), type = "lines", col = "orange")

lines(drmPred1$PrDens~drmPred1$calBP, col = "black")
lines(drmPred2$PrDens~drmPred2$calBP, col = "blue")
lines(drmPred3$PrDens~drmPred3$calBP, col = "orange")
}

```

```

FromTo <- 1:clusterCounts[i]

startBuilder <- c("starter <- c(
      paste("drm", FromTo, "$coefficients," , sep = ""),
      "")")

startBuilder[(length(startBuilder)-1)] <- gsub("$", "",
startBuilder[(length(startBuilder)-1)])

eval(parse(text = startBuilder))

names(starter) <- gsub(":.*", "", names(starter))

unchangedStarter <- starter

starter2 <- starter[which(names(starter) == "d")]
starter <- starter[-which(names(starter) == "d")]

starter1 <- data.frame(names = names(starter), value = unname(starter))

model <- buildModel(clusterCounts[i], "calBP", "PrDens", starter = starter1)

# model

```



```

nlsFinal <- nls(model, data = SPD, start = starter2, upper = (starter2), lower =
(starter2), algorithm = "port")
plot(SPD$PrDens~SPD$calBP, type = "lines") +
  lines(nlsFinal$m$fitted()~SPD$calBP, col = "blue")
modelList[[i]] <- nlsFinal
paramList[[i]] <- unchangedStarter

predGrid[[i]] <- data.frame(calBP = abs(SPD$calBP), PrDens = nlsFinal$m$fitted())
robustKeeper <- as.numeric(robust1[which(methodSelection ==
min(methodSelection)),])
robustList[i,c(1:length(robustKeeper))] <- robustKeeper
}

}

BIC(modelList[[i]])+(log(nrow(SPD))*3*clusterCounts[i])

}

for(i in 1:length(modelList)){

  BICdf[i,] <- cbind(Unit = fileNumbers[i],BIC =
BIC(modelList[[i]])+(log(nrow(SPD))*3*clusterCounts[i]))

}

writeLines(modelsExport, paste(csvOutputs, "Models.txt", sep = ""))

write.csv(BICdf, paste(csvOutputs, "BICmultilog.csv", sep = ""), row.names = F)

```

```

SPD <- read.csv(SPDLList[i])
plot(SPDS$PrDens~SPD$calBP, type = "lines") +
  lines(modelList[[i]]$m$fitted()~SPD$calBP, col = "blue")

mTList <- list()

for(i in 1:length(fileNumbers)){

  eval(parse(text = paste("Sbox <- SboxList[["",
    fileNumbers[i],"\"]",
    sep = "")))

  cptcal <- calibrate(x = Sbox$Age,
    errors = Sbox$Sd,
    normalised = F,
    ncores = 6)

  predGrid[[i]]$calBP <- abs(predGrid[[i]]$calBP)
  mTList[[i]] <- modelTest(cptcal,
    errors = Sbox$Sd,
    nsim = 1000,
    bins = NA,
    runm = 250,
    timeRange = c(abs(min(SPDS$calBP)),
      abs(max(SPDS$calBP))),
    model = "custom",
    normalised = F,
    raw = T,
    edgeSize = -550,
    ncores = 6,

```

```

        verbose = T,
        predgrid = data.frame(predGrid[[i]])
mT <- mTList[[i]]$result
mT$scalBP <- -abs(mT$scalBP)
predGrid[[i]]$scalBP <- -abs(predGrid[[i]]$scalBP)
png(file = paste(currentPath,
        "Images/Unit ",
        Sbox$Sbox[1],
        ifelse(i == 4,
                "MultiLogistic",
                "Bilogistic"),
        ".png",
        sep = ""),
height = 8.5,
width = 11,
units = "in",
res = 300,
family = "Times")
plot(mT$PrDens ~ mT$scalBP,
type = "lines",
main = paste(boxLookup[which(boxLookup[,1] %in% Sbox$Sbox[1]),2],
        ifelse(i == 4,
                " Multilogistic",
                " Bilogistic"),
        " Curve With Envelope",
        sep = ""),
ylim = c(min(min(mT$PrDens),
        min(mT$lo)),

```

```

        max(max(mT$PrDens),
            max(mT$hi)),
    xlab = "calBP",
    ylab = "PrDens")
polygon(c(mT$scalBP, rev(mT$scalBP)),
        c(mT$lo, rev(mT$hi)),
        col = "grey",
        lty = 0)
lines(mT$PrDens~mT$scalBP)
lines(predGrid[[i]]$PrDens~predGrid[[i]]$scalBP,
      col = "blue")
dev.off()
pdf(paste(pdfOutputs,
          "Unit ",
          Sbox$Sbox[1],
          ifelse(i == 4,
                "MultiLogistic",
                "Bilogistic"),
          ".pdf",
          sep = ""),
    height = 8.5,
    width = 11)
plot(mT$PrDens ~ mT$scalBP,
     type = "lines",
     main = paste(boxLookup[which(boxLookup[,1] %in% Sbox$Sbox[1]),2],
                 ifelse(i == 4,
                       " Multilogistic",
                       " Bilogistic"),

```

```

      " Curve With Envelope",
      sep = ""),
ylim = c(min(min(mT$PrDens),
              min(mT$lo)),
          max(max(mT$PrDens),
              max(mT$hi))),
xlab = "calBP",
ylab = "PrDens")
polygon(c(mT$calBP, rev(mT$calBP)),
        c(mT$lo, rev(mT$hi)),
        col = "grey",
        lty = 0)
lines(mT$PrDens~mT$calBP)
lines(predGrid[[i]]$PrDens~predGrid[[i]]$calBP,
      col = "blue")
dev.off()
residuals <- nlsFinal$m$resid()
laggedresiduals <- lag(residuals)
png(file = paste(currentPath,
                 "Images/Unit ",
                 Sbox$Sbox[1],
                 "resid~lagged(multilog).png",
                 sep = ""),
    height = 8.5,
    width = 11,
    units = "in",
    res = 300,
    family = "Times")

```

```

plot(residuals~laggedresiduals)
dev.off()

}

for(i in 1:4){
  SPD <- read.csv(SPDList[i], header = T)

  inflection1 <- (paramList[[i]][which(paramList[[i]]$names == "limit1"),2] -
unnamed(modelList[[i]]$m$getPars())/2

  plot(SPD$PrDens~SPD$calBP, type = "lines") +
  abline(h = inflection1)

  when <- SPD[which(min(abs(SPD$PrDens-inflection1)) == abs(SPD$PrDens-
inflection1)),c(1,2)]

  inflection2 <- (paramList[[i]][which(paramList[[i]]$names == "limit1"),2] -
unnamed(modelList[[i]]$m$getPars()+paramList[[i]][which(paramList[[i]]$names ==
"limit2"),2])/2

  when2 <- SPD[which(min(abs(SPD$PrDens-inflection2)) == abs(SPD$PrDens-
inflection2)),c(1,2)]

  inflection3 <- (paramList[[i]][which(paramList[[i]]$names == "limit1"),2] -
unnamed(modelList[[i]]$m$getPars()+paramList[[i]][which(paramList[[i]]$names ==
"limit2"),2]+paramList[[i]][which(paramList[[i]]$names == "limit3"),2])/2

  when3 <- SPD[which(min(abs(SPD$PrDens-inflection3)) == abs(SPD$PrDens-
inflection3)),c(1,2)]

  cat(paste("\n\n",when[,1], when[,2], when2[,1], when2[,2], when3[,1], when3[,2]))
}

```