

# Critique of Laufenberg Paper

## Stat 512:Final Project

Steve Huysman & Parker Levinson

Friday, May 5, 2023

## Contents

0.1	Introduction	.	.	.	.	.	.	.	.	.	.	1
0.2	Methods	.	.	.	.	.	.	.	.	.	.	2
0.3	Results/Summary of Statistical Findings	.	.	.	.	.	.	.	.	.	.	9
0.4	Scope of Inference	.	.	.	.	.	.	.	.	.	.	10
0.5	Critique	.	.	.	.	.	.	.	.	.	.	10
0.6	Group Work Statement	.	.	.	.	.	.	.	.	.	.	10
0.7	References	.	.	.	.	.	.	.	.	.	.	11
0.8	Appendix	.	.	.	.	.	.	.	.	.	.	11

## 0.1 Introduction

Whitebark pine (*Pinus albicaulis* Engelm.; abbreviation: WBP) is a conifer tree native to the mountains of the western United States and Canada. It inhabits subalpine areas where it can be found growing up to the tree line, at a higher elevation than other tree species found with it. Whitebark pine is an early successional species that is often the first to establish after disturbance such as wildfire. WBP is a keystone species of subalpine environments where it plays important ecological roles such as providing food for wildlife such as Clark's Nutcracker and the threatened Grizzly Bear.

Due to threats from climate change, mountain pine beetle, and the invasive white pine blister rust, Whitebark pine has undergone a rapid and widespread decline. It was recently estimated that over half of all standing WBP in the United States are dead. This decline has lead to its recent listing as Threatened under the Endangered Species Act. Future climate

projections indicate further deterioration of WBP's habitat. Strategies to conserve this species involve planting WBP seedlings for restoration of high-elevation forests. Successful plantings in the face of climate change require an understanding of the relationship between climate and seedling establishment and growth in this species. Competition from other tree species also plays a role in seedling establishment and was investigated here.

Laufenberg et al. (2020) investigated USFS plantings of WBP seedlings in the Greater Yellowstone Ecosystem (GYE) to answer two research questions:

- 1) What is the relationship between climate/competition and WBP seedling establishment, measured by individual growth rate?
- 2) What is the relationship between climate/competition and WBP seedling survival, measured by density change?

However, we chose to focus on a simplified question. We are interested in how annual evapotranspiration and water availability affect planted WBP seedling growth after accounting for location, competition, and climatic variables.

## 0.2 Methods

### 0.2.1 Field Methods/Study design

Over the past 40 years, the US Forest Service and National Park Service has planted more than 1500 acres of WBP in the GYE. This study investigated five planting units (Figure 1) that each contained between two and eight planting sites. This study used a hierarchical sampling design including 5 planting units, with a total of 29 planting sites across units, and thousands of white bark pine seedlings per planting site. (See Figure 2 for model design.)

In the 5 larger planting units, no data aside from name of the unit, were collected. For the 29 sites scattered across those planting units (an unbalanced design), larger climatic data were recorded such as average annual temperature and average annual precipitation. (For a full list of data collected at each location, see Figure 2. Then, seedlings within that site were sampled for annual growth rate. Seedlings were too small to measure growth rings via coring,

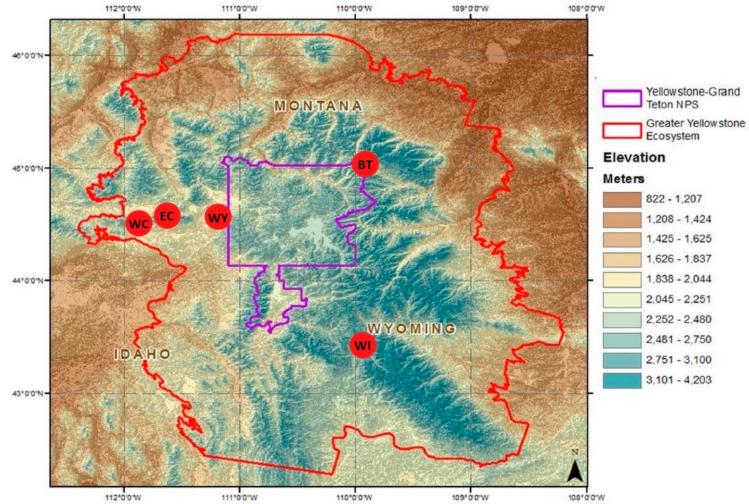


Figure 1: Map of planting units included in study in the Greater Yellowstone Ecosystem. Yellowstone National Park is outlined in purple, the greater yellowstone ecosystem is outlined in red, and the five white bark pine planting units run by the Forest Service and National Park service are circled in red. The two-letter abbreviation denotes the name: BT = Beartooth, EC = East Centennial, WC = West Centennial, WI =Wind River, WY =West Yellowstone

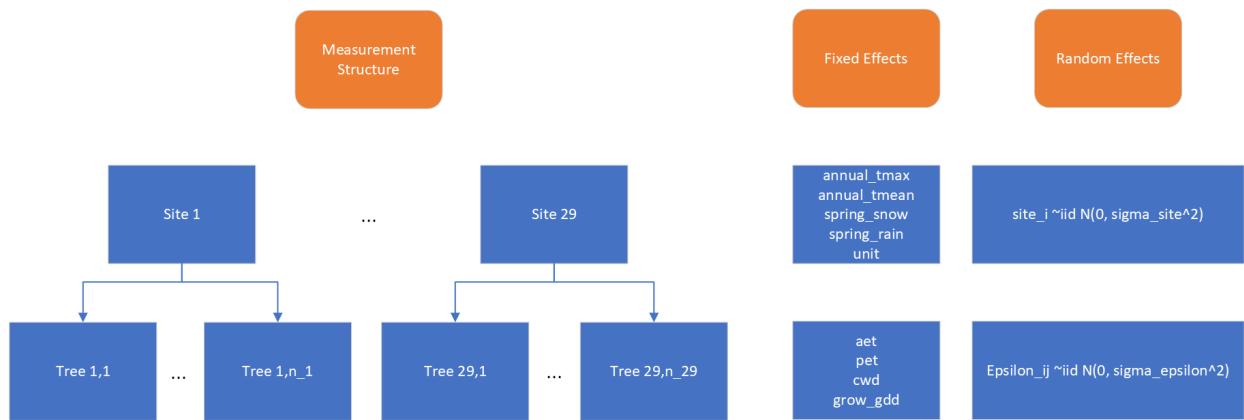


Figure 2: Hierarchical study design with fixed and random effects.

so height was used as a proxy for growth rate. Specifically, growth rate was calculated as the change in height between the study year (2018) to the relative planted height when the seedling was first planted. This was divided by the number of years since planting minus 2.5 years to account for the period of time when seedlings sequester carbon instead of focusing on their own growth.

At each site, climate data like elevation, aspect, water deficit, number of competitors, and amount of evapotranspiration were collected. Water balance variables were estimated using Daymet devices as inputs for temperature and precipitation. Then, specific data were taken for each individual tree, including height, number of competitors, and climatic variables. Sampling individual seedling growth occurred from May 2018 to October 2018. A grid cell matrix of 10 meters x 10 meters was overlaid on the study site or unit. A random starting point was decided and then every 20th grid cell from that was sampled, equating to sampling WBP in 2-15% of each site. Each seedling within that grid cell was digitally tagged, and Survey123 was used to collect field data. Researchers were most interested in the annual growth rate. Seedlings were too small to measure growth rings via coring, so height was used as a proxy for growth rate. Specifically, growth rate was calculated as the change in height between the study year (2018) to the relative planted height when the seedling was first planted. This was divided by the number of years since planting minus 2.5 years to account for the period of time when seedlings sequester carbon instead of focusing on their own growth.

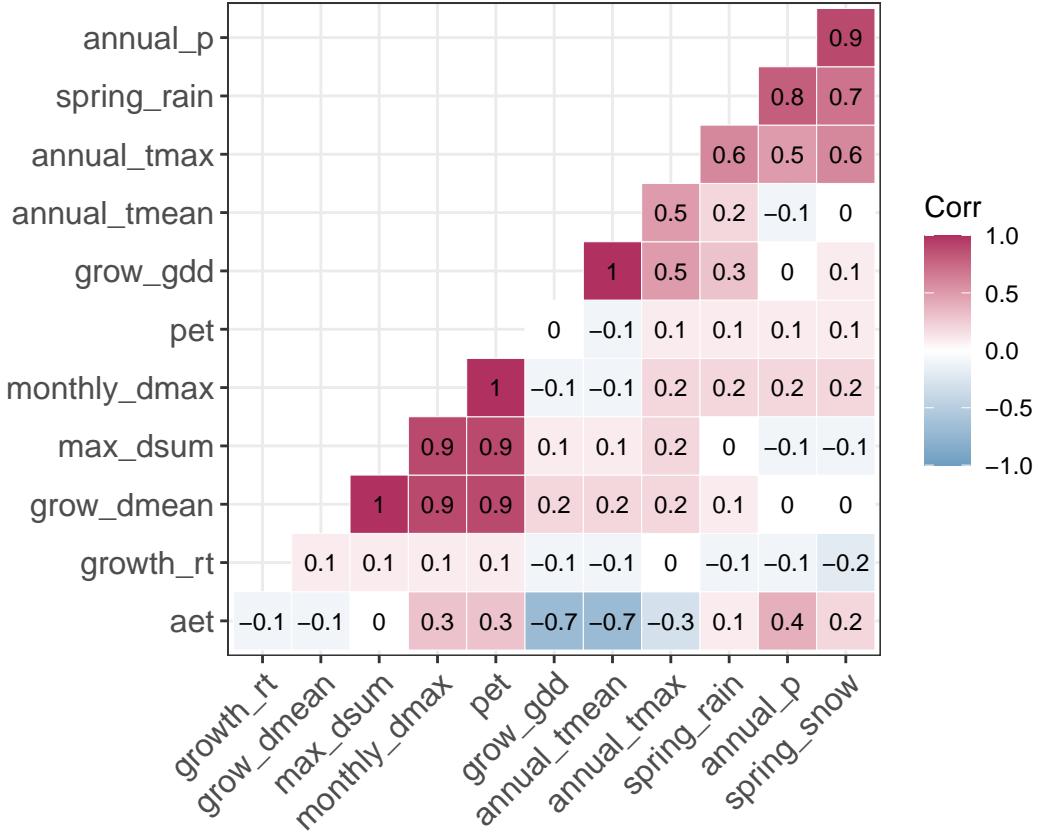
A variety of data were collected at different scales. The variables that are most pertinent to the research question are:

- WBP seedling height: This is a continuous response variable and is used as a proxy for seedling growth.
- AET: This is a continuous predictor variable for annual evapotranspiration, which indicates the usable amount of energy in the system available for seedlings to use. This is collected at each seedling.
- CWD: This is a continuous predictor variable for climatic water deficit, which indicates how much more water a plant could use. This is col-

lected at each seedling. - Micro - This was a binary variable indicating presence of favorable microsite conditions. 1 if there was a rock or other topographical feature that changed the environmental conditions where the seedling lived.

Data were provided by Laufenberg, but the process of cleaning and structuring the data were not explicit in the paper. As such, we had to experiment to figure out how data were cleaned. We ultimately decided to create an annual evapotranspiration rate, given that it is extremely pertinent to our research question. To get this annual evapotranspiration (AET), we multiplied the average monthly evapotranspiration by 7 for the 7 months of the growing season. The same was done for mean water deficit (grow\_dmean) to get climatic water deficit (CWD).

Climate and water balance predictor variables were tested for colinearity with a cutoff of  $r \geq +/- 0.6$  (Figure ??). A parsimonious list of water balance variables was selected by choosing the more biologically relevant variable from pairs that exceeded this threshold. This list of variables was then combined with the variables pertinent to the research question, which are *AET* and *WD*. Following this process, our list of environmental components we would like to account for in the model are comp\_number and microsite status.



After data were selected based on their correlation, we looked at linearity and potential other polynomial forms (Figure 3). From examination of the raw data, there is a little variance in growth rate by planting unit, which we will account for with a fixed effect of unit.

From examination of the raw data, growth rate appears to be similar across each planting unit, suggesting that no transformation is needed (Figure 4).

### 0.2.2 Statistical Procedures Used

Mixed effects models were used to study what covariates were correlated with WBP performance. Random effects were always used for the site and fixed units were used for the study unit, due to only having 5 units. A variety of linear and polynomial functions, and fixed effects were explored. Cubic forms were explored due to the biological idea that WBP seedling establishment may vary drastically depending on the conditions. In below average

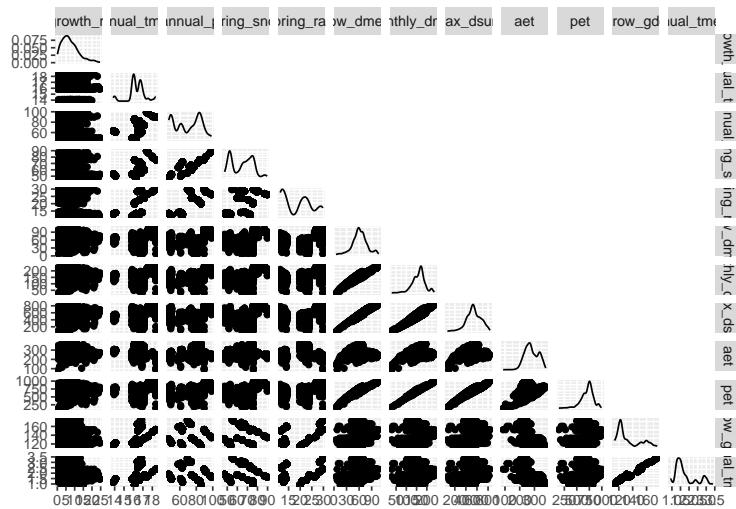


Figure 3: Raw data visualization of variables of interest

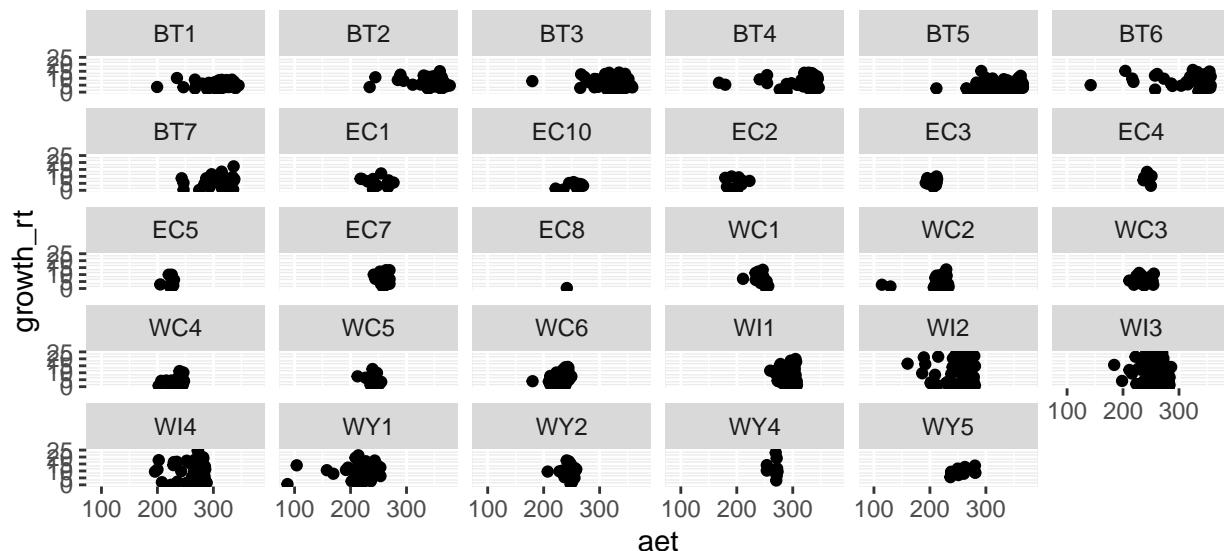


Figure 4: Growth rate compared with AET by each site

conditions, we expect low growth rate. We expect there to be an average growth rate in average conditions, and then for high quality conditions, we expect a much higher growth rate - therefore following a cubic form. A corrected AIC was used to compare models. We aimed to find the most parsimonious set of water balance variables while still incorporating the biologically relevant variables and incorporating environmental variables related to the research questions (AET and CWD).

The theoretical model for our full model analyzed is as follows:

$$\log(growth\_rate)_{ij} = \mu_{ij} + Site_i + \epsilon_{ij} \quad (1)$$

$$Unit_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{Site}^2) \quad (2)$$

$$\epsilon_{ij} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{error}^2) \quad (3)$$

$$\begin{aligned} \mu_{ij} = & \beta_0 + \beta_1 * AET + \beta_2 * CWD + \beta_3 * I_{Unit=WC} + \beta_4 * I_{Unit=EC} \\ & + \beta_5 * I_{Unit=WI} + \beta_6 * I_{Unit=BT} + \dots \beta_i * climatevariables \end{aligned} \quad (4)$$

$$I_{Unit=WC} = \begin{cases} 1, & \text{if } unit = WestCentennial \\ 0, & \text{otherwise} \end{cases} \quad I_{Unit=EC} = \begin{cases} 1, & \text{if } unit = EastCentennial \\ 0, & \text{otherwise} \end{cases} \quad I_{Unit=WI} = \begin{cases} 1, & \text{if } unit = WI \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

We ran the saturated cubic model without interactions (Equation (4)), and then refined it using backwards selection from the *step()* function and a AIC cutoff of AIC<2. After running the model initially, we found that the residuals had a lot of funneling. Because of this, we chose to log transform the response variable (*growth\_rate*) to reduce this skew. However, log transforming does result in a non-random distribution of residuals. All analysis was done use R statistical software (v4.2.2; R Core Team 2021).

### 0.3 Results/Summary of Statistical Findings

Using backwards selection, the most parsimonious model is below, where *annual\_Tmax* took on a quadratic form, and PIEN and Micro were removed. This model has an AIC score of 1111.89.

$$\begin{aligned} \log(growth\_rate)_{ij} &= \hat{\mu}_{ij} + Unit_i + \epsilon_{ij} \\ Unit_i &\stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{unit}^2) \\ \epsilon_{ij} &\stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{error}^2) \\ \hat{\mu}_{ij} &= \beta_0 + \beta_1 AET^3 + \beta_2 AET^2 + \beta_3 AET \\ &+ \beta_4 PET^3 + \beta_5 PET^2 + \beta_6 PET \\ &+ \beta_7 Comp\_number^3 + \beta_8 Comp\_number^2 + \beta_9 Comp\_number \\ &+ \beta_{10} PPT^3 + \beta_{11} PPT^2 + \beta_{12} PPT \\ &+ \beta_{13} T_{max}^2 + \beta_{14} T_{max} + \beta_{15} PICO + \beta_{16} ABLA \end{aligned}$$

The initial full model, with all variables of interest, had an AICc of 1126.5. This was a 14.62 reduction in AICc value.

The final model suggests the WBP seedling growth is most correlated with the annual evapotranspiration, the potential evapotranspiration, the number of competitors around the seedling, annual precipitation, and max temperature while accounting for the random effect of planting unit.

Going forward, we would like to look into the following things: - We get the error “fixed-effect model matrix is rank deficient so dropping 14 columns / coefficients” and need to look at what that means.

- Better understand the variable selection process and why certain variables were included.
- Look for a more parsimonious model, perhaps dropping cubic forms.
- Creating diagnostic plots for mixed models - Interpreting the coefficients and the 95% CI

in the context of WPB restoration

- Insert model table - Effects plot/model estimate - Random effect for sample site? - Better define initial model and which betas are important

## 0.4 Scope of Inference

Because there was random sampling of the WPB within each site, the inferences can be applied to the larger WBP population at that site, but because the sites and units were not randomly selected, no further inferences can be drawn. Furthermore, there was no treatment applied since this was an observational study, so inferences are only correlative, not causative.

## 0.5 Critique

Trying to follow and reproduce what Laufenberg et al. 2020 did in the paper proved to be extremely complicated. There appeared to have been extensive data cleaning and manipulation, which was not communicated in the paper. We know that they chose the “more ecologically relevant” variable when looking at correlation, but there was no clear instructions on what that was. Furthermore, the paper had no clear model selection processes, making it difficult to reproduce. Many of their choices, such as using a cubic form for AET or log transforming growth rate, were not explained and the reasoning behind was unclear. The paper itself is fishing for an explanation.

After trying and failing to reproduce the results from the paper, we streamline our research question and attempted to just use the data collected to answer what we were interested in. However, even that proved to be challenging.

## 0.6 Group Work Statement

Steve spearheaded the data wrangling and analysis, while Parker took lead on the write-up and formatting. Both people contributed to the model selection, interpretation, discussion,

and critique.

## 0.7 References

- Laufenberg, David, et al. “Biophysical gradients and performance of whitebark pine plantings in the Greater Yellowstone Ecosystem.” *Forests* 11.1 (2020): 119.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi: 10.18637/jss.v067.i01.

## 0.8 Appendix

- Must include a compiled RMarkdown with all of our results
- All code is in currently in an R script, will attach for final paper