Eileen Arata

**Q SCI 483 Final:**

**Assessing Forest Restoration using Seed Counts**

**Abstract**

Forest restoration is an essential tool for managing damaged forests in the Pacific Northwest. The Nature Conservancy is taking a landscape-scale approach to restoration in their Ellsworth Creek Preserve in coastal Washington. To assess the effect of restoration practices, the productivity of treated and untreated stands is being evaluated with conifer seed traps. This preliminary study of the seed trap contents explores a variety of regression models with a range of predictors to find which attributes affect productivity the most and whether or not restoration treatments have had a significant effect. In the end, a negative binomial model with stand age, stand aspect, and treatment as predictors was found to be the best fit to the data. Age, aspect, and treatment all had positive influences on seed counts indicating that treatments are having an effect, but stand characteristics also significantly influence the productivity of a forest stand.

1. **Introduction**

Forest restoration is a complex task with numerous management avenues, but it has become increasingly important with recent natural disasters and goals set by the Northwest Forest Plan (Thomas et al., 2006). That is why evaluating the success of different kinds of restoration is so essential to build knowledge of the best practices and increase its effectiveness (Crouzeilles et al. 2016). The use of seed traps to quantify tree productivity after restoration or disturbances has been found extremely effective, despite significant amounts of manual effort (Tattoni et al., 2021). By using this method, conifer productivity can be used to evaluate influences on forest restoration and its success.

**Figure 1** Location of the Ellsworth Creek Preserve in relation to Seattle. The preserve is denoted by the red pointer.

Map

Description automatically generated The data for this study comes from the Ellsworth Creek Preserve on the coast of Washington State. (**Figure 1**), which is managed by The Nature Conservancy (TNC). After purchasing the whole 7,600 acres from various landowners starting in 1998, TNC implemented both active and passive restoration techniques at a landscape level. This land ranged from pure clearcuts to quasi-monocultures to a fully intact stand of old growth forest. With this variability and different restoration techniques, TNC is using this unique landscape to study the effects of forest restoration on wildlife, local hydrology, and conifer productivity.

The scope of this paper is evaluating conifer productivity throughout various stands with different attributes in the Ellsworth Preserve using seed trap data. Pilot studies and previous work in the preserve have indicated that age is a significant predictor of how many seeds are in a stand, with older stands having more productivity. Additionally, stands where thinning was conducting are hypothesized to have higher seed counts since fewer trees in the stand should exert less competitive pressure, allowing the remaining trees to produce more seeds. Other predictors are hypothesized to have a negligible effect on seed counts.

1. **Methods**
   1. **Data Description**

This project is part of a larger study conducted by TNC Quantitative Ecologist Ailene Ettinger. While the full study will use samples from multiple time periods and more stands, this paper will focus on samples collected from a single 6-month period from November 2021 to May 2021. At the beginning of the study period, seed traps built of laundry baskets with netting were set up within randomly chosen stands and subplots (north, east, south, and west) throughout the preserve. Cones and seeds from the conifers above the baskets would fall into the netting, which are collected into bags for processing at the end of the study period. Processing these samples involves going through the contents of each basket, finding seeds and cones, and classifying these by species. At the time of this analysis, 44 of these samples have been processed, representing 13 stands and 44 subplots.

The seed count data was then combined with other stand-level information, such as age, treatment status, slope, and elevation. The attributes analyzed are shown in **Table 1**. The total count of seeds found in each subplot is represented as “Seed Count” and is the response variable in all analyses. Stand age is represented as “Age,” meaning the average age of the stand estimated by TNC and measured in years. Treatments were split into two categories: control, which received no restoration treatment, and thinned. The thinned category encompasses all types of thinning used. Together, this information sufficiently describes the stands being studied and attributes that may be pertinent to tree productivity.

**Table 1** Description of data attributes used in analyses.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Attribute** | **Units** | **Data Range** | **Variable Type** | **Amount** |
| Seed Count | seeds | Discrete: 4-924 | Response | 44 samples |
|  |  |  |  |  |
| Stand Age | years | Discrete: 15-85 | Predictor | 13 stands |
| Treatment |  | Factor: control or thinned | Predictor | 22 samples thinned |
| Aspect | degrees | Continuous: 0-360 | Predictor |  |
| Slope | degrees | Continuous: 26-94 | Predictor |  |
| Elevation | feet | Continuous: 60-282 | Predictor |  |
| Stand |  | Factor: stand name  (i.e., C2-13) | Random Effect | 13 stands |
| Subplot |  | Factor: N, E, S, W | Random Effect | 4 subplots/stand  (subplots with no data were removed, so some stands do not have all 4) |

* 1. **Data Manipulation**

All subsequent analyses were run in R (version 4.2.2) using R Studio. The packages used were dplyr, car, MuMIn, MASS, and AER.

Using forward stepwise model construction, a series of simple linear models were built to assess which predictors would be the most valuable to explore. This mechanism starts with simple linear models with one predictor, then adds predictors to the best model from each round. The cycle is completed when the AICc score for the best model stops decreasing when new predictors are added because this indicates that the new predictors are just adding complexity and no explanatory benefit to the model. As shown in **Table 2**, AICc scores stopped decreasing after models with two predictors, and the most parsimonious linear model had age and aspect as predictors. This model was then used to assess how the data interacted with a regression model, such as influential points and potential transformations.

**Table 2** The results of initial simple linear models with 1-3 predictors ordered by AICc score where the lowest score is the most parsimonious model. The adjusted R2 value indicates how much variation (out of 1.0) is explained by the model. The adjusted value is used because it only increases if the added predictor has a non-negligible effect.

|  |  |  |
| --- | --- | --- |
| **Simple Linear Models** | | |
| **Model** | **AICc** | **Adjusted R2** |
| **Age + Aspect** | 573.7492 | 0.1066 |
| **Age** | 574.3951 | 0.06484 |
| **Aspect** | 575.7541 | 0.0355 |
| **Null** | 576.0727 | 162.4 residual SE on 43 DF |
| **Age + Aspect + Slope** | 576.2776 | 0.08481 |
| **Age + Slope** | 576.7644 | 0.04326 |
| **Age + Elevation** | 576.7761 | 0.043 |
| **Age + Treatment** | 576.8178 | 0.04209 |
| **Elevation** | 577.7195 | -0.008556 |
| **Treatment** | 578.3542 | -0.02321 |
| **Slope** | 578.3765 | -0.02373 |
| **Age + Aspect + Treatment** | 579.1858 | 0.02227 |
| **Age + Aspect + Elevation** | 579.2514 | 0.02081 |

Outliers and influential points were determined using Cook’s Distance and DFFITS. Only one point, Stand C1-18 Subplot E Count 924, was considered an outlier and influential by both these metrics. Its Cook’s Distance was 0.748, which was well over three times the average Cook’s Distance. Additionally, its DFFITS value was far above one critical value, meaning that it is influential as well as an outlier. The point is not removed in subsequent models since it was a true seed count, not a recording error. However, it is taken into consideration as influential when assessing later model trends and diagnostics.

Chart, scatter chart

Description automatically generated When evaluating potential predictor transformations, neither age nor aspect needed to be transformed since the points were generally well-distributed along the x-axis for both. The response, on the other hand, required transformation because of the influential point discussed earlier. The boxcox() function in the MASS package was applied to the age + aspect model, which indicated that a log transformation of the response would be the most appropriate (**Figure 2**). The log transformation was then applied to all subsequent linear models. but not generalized linear models.

**Figure 2** Illustrating the transformation of the response variable, seed counts. Only one predictor, age, is shown for ease of graphing and visibility, but aspect was included in the linear models shown by the red lines.

Since the two predictors, age and aspect, were initially found to explain seed counts, they were checked for multicollinearity to ensure that their effects were not muddied within models. With a Pearson’s correlation coefficient of -0.0251 and a pair’s plot that showed no patterns, these predictors were not found to be correlated. Due to this lack of correlation, no interaction effects were included in models with both these predictors.

* 1. **Model Building & Analyses**

Models were built with increasing complexity to evaluate the effects of all predictors on seed counts. Based on the initial findings from the simple linear models and the results of the dredge() method, age and aspect were again found to be the most influential predictors when the response was log transformed. The best multiple regression model had the equation:

log(counti) = 1.971 + 0.0313(agei) + 0.00258(aspect) + εi

The parameter estimate for age, 0.0313 (95% CI: 0.0170 to 0.0456) was found to be significant with a p-value of 7.26e-5, which is less than ⍺ = 0.05, indicating that it has a nonzero effect on seed count. The aspect parameter estimate of 0.0026 (95% CI: -0.0006 to 0.0058) was not found to be a significant predictor with a p-value of 0.109. Additionally, the assumptions for multiple regressions were all violated with diagnostic plots showing clear issues with linearity, constant residual variance, and normally distributed residuals in the first half of the fitted values. Even with the transformed response, the best multiple regression linear model did not fit the data well, indicating that the data does not follow a strictly linear pattern.

After the linear model did not fit the data well, polynomial terms were explored for both the age and aspect predictors. The dredge() function was run for all combinations of linear and second-order polynomial terms for both age and aspect. The age2 term appeared in the top 8 models while age, aspect, and aspect2 appeared 4 times apiece. The model with the lowest AICc score included both the age and age2 terms, shown in this equation:

log(counti) = 4.749 – 0.0889(agei) + 0.00122(agei)2 + εi

The parameter estimate for age is 0.0889 (95% CI -0.176 to -0.00195), and the estimate for age2 is 0.00122 (95% CI: 0.000346 to 0.00209). The adjusted R2 value is 0.389, which is higher than any linear model. Even though this model is nonlinear and the residuals are not normally distributed on the extremes of the QQ plot, the variance of the residuals is far more homogenous than the simple linear models. Despite the improved diagnostics and explained variation, the polynomial model was not found to fit the data as well as generalized linear models that were able to account for the response variable being only positive and with discrete values.

Because the response is a discrete count variable that must be positive, the Poisson distribution was used in the next model, a generalized linear model. The dredge() function was used on a glm with the Poisson distribution and log link function using the untransformed seed counts as the response and all the linear predictor variables. The model with the lowest AICc score included all of the predictors except elevation where:

log(counti) = 0.705 + 0.0221(agei) + 0.00420(aspecti) + 0.0200(slope) + 2.160(treatTHINi)

counti ~ Poisson(mean(count)i)

Since almost all of the predictors were included, this model is not particularly useful for singling out individual influences on seed counts. Additionally, model diagnostics indicate that this model, despite being a more appropriate design for counts, does not fit the data well. A χ2 goodness of fit test on the residual deviance resulted in a p-value of 0, meaning that the null hypothesis that this model is a good fit was rejected. The Poisson assumption of mean = variance was also violated according to a dispersion test where the p-value of 0.0439 was less than ⍺ = 0.05, indicating that the true dispersion is greater than 1. Cook’s distance also found 6 influential points skewing the data. Nevertheless, accounting for a discrete response variable is worthwhile for this dataset.

Considering this, the negative binomial distribution was applied for the best and final model. The forward stepwise model selection method was used, as described above, until the model could not iterate to completion due to too many predictors. The model with the lowest AICc score that resulted from this method was:

log(mean(count)i) = 2.339 + 0.0213(agei) + 0.00422(aspecti) + 0.688(treat\_THINi)

θ = 0.998

counti ~ Negative Binomial(mean(count)i, θ)

The parameter estimate for age2 is 0.0213 (95% CI: 0.00928 to 0.0328), and the parameter estimate for aspect is 0.00422 (95% CI: 0.00141 to 0.00698). Locations where thinning was applied included the treatment intercept estimate of 0.688(95% CI: 0.337 to 1.315). This model fit the data far better than its Poisson counterpart since its AICc score was 4683.472 lower and was consequently chosen as the best model for this data.

1. **Results**

**Figure 3** The relationship between the modelled mean and variance for the Poisson and negative binomial models.

Chart

Description automatically generatedThe negative binomial model was chosen for its appropriate handling of the count data and improved performance over the Poisson model. The low θ value of 0.998 indicates that it handles the variable relationship between the variance and mean better than the Poisson (**Figure 3**). Additionally, the residual deviance of 50.356 on 40 degrees of freedom indicates that much of the variation in the data is explained by the model. To quantitatively emphasize the fit, the χ2 goodness of fit test on the deviance had a p-value of 0.126, which means that we fail to reject the null hypothesis of the model being a good fit. Taken together, this distribution, predictors, and diagnostics all indicate that this negative binomial model is the best fit for the seed count data.

The results of the best model in the original units for the data are shown in **Table 3**.

**Table 3** The results of the negative binomial model used in this analysis. The equation is presented in link units with the log link function. The parameter values are presented in original scale units.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Equation:** | log(mean(count)i) = 2.339 + 0.0213(agei) + 0.00422(aspecti) + 0.688(treat\_THINi) | | | |
| **Link Function:** | log(mean(count)) | | | |
| **Distribution:** | counti ~ Negative Binomial(mean(count)i, 0.998) | | | |
| **Parameters:** | **Estimate**  **(original scale)** | **Lower 95% CI**  **(original scale)** | **Upper 95% CI**  **(original scale)** | **P-value**  **(link scale)**  **(\* = significant at ⍺= 0.05)** |
| Intercept  (treatCON) | 10.372 seeds | 3.730 | 35.002 | 4.59e-5\* |
| Age | 1.022 seeds/yr | 1.009 | 1.033 | 6.97e-4\* |
| Aspect | 1.004 seeds/degree | 1.001 | 1.007 | 4.74e-3\* |
| Intercept  treatTHIN | 12.363 seeds | 11.407 | 14.098 | 0.0384\* |

Based on the negative binomial model, age, aspect, and treatment all have a positive, significant relationship with seed count. For each year of stand age, its seed count will increase by 1.022 seeds/year (95% CI: 1.009 seeds/year to 1.033 seeds/year). For each degree of aspect for the stand position, its seed count will increase by 1.004 seeds/degree (95% CI: 1.004 seeds/degree to 1.007 seeds/degree). This indicates that stands at higher degrees, particularly facing west, have higher seed counts. Finally, stands that are treated have a higher intercept than those that are not by 0.0337 seeds (95% CI: 1.034 seeds to 3.726 seeds). The higher intercept for thinned stands is significant at the ⍺ = 0.05 level, meaning that the rate of change for those stands in regards to the other parameters is the same, but the count of seeds is higher.

**Figure 3** shows the fitted model values for both controlled and thinned plots using age and aspect as predictors. For both attributes, seed counts increase as their respective values increase, as illustrated by the positive curves of the fitted lines. However, the width of the 95% confidence intervals also increase at higher values, particularly for thinned plots, meaning there is more uncertainty around higher fitted values.

**Figure 3** Graphing the fitted values produced by the negative binomial model for both controlled and thinned stands. Aspect and age are held constant at their respective means for the other predictor’s graph. The solid lines are the fitted values, and the faded dotted lines are 95% confidence intervals.

Chart

Description automatically generated

1. **Discussion/Conclusions**

This introductory analysis of seed trap data from the Ellsworth Preserve has shown that age and aspect are important predictors for seed count, and, by extension, the productivity of a stand. Older stands are more likely to have larger trees since they have had more time to grow, which could in turn mean more branches producing cones and, therefore, seeds. While older stands are more productive, as expected, it is encouraging to see that the best model found restoration treatments increased seed counts as well. Thinning treatments may have diminished competition among different species in these stands, allowing the remaining trees to spend more energy on reproduction. As the treated stands grow older, both of these effects will combine to maximize seeds production and hopefully result in resilient, self-sufficient stands.

Aspect was an unforeseen significant predictor for seed counts, but not altogether unsurprising. Ellsworth is located directly on the coast of western Washington, and western facing stands, which would have higher aspect degrees between 180° and 360°, may receive more moisture from the ocean. The findings of this model may be the result of a localized rain shadow effect. Without having to mitigate moisture levels, trees in western facing stands might be able to allocate more resources into producing cones.

Even though these findings highlight emerging trends in the data, it is important to note that this analysis was not done on the full dataset. Seed trap contents are still being processed from multiple time periods and stands. An extended analysis on all the data could either confirm these trends or reveal entirely different ones. Additionally, this paper focused on regression models for the purposes of being a final course project. A more complete study could include a variety of other models, such as ANOVAs for categorical variables. Another drawback to note is the influence of certain outlying datapoints. The main outlier came from Stand C1-18 Subplot E with a seed count of 924, which was much higher than any other stand. This could be due to a conifer tree directly above the seed trap or wind patterns or an entirely different factor. Either way, it made the variances of models difficult to quantify.

Despite these limitations, this preliminary analysis of seed trap contents from the Ellsworth Preserve found that older, western facing stands that have been thinned are significantly more likely to produce more seeds.

**Works Cited**

Crouzeilles, Renato, Michael Curran, Mariana S. Ferreira, David B. Lindenmayer, Carlos E. Grelle, and José M. Rey Benayas. “A Global Meta-Analysis on the Ecological Drivers of Forest Restoration Success.” *Nature Communications* 7, no. 1 (2016). <https://doi.org/10.1038/ncomms11666>.

“Ellsworth Creek.” The Nature Conservancy. Accessed June 2, 2023. <https://www.nature.org/en-us/get-involved/how-to-help/places-we-protect/ellsworth-creek/>.

Minore, Don. “Germination, Survival and Early Growth of Conifer Seedlings in Two Habitat Types.” *USDA Forest Service Research Paper*, January 1986. <https://doi.org/10.2737/pnw-rp-348>.

Stewart, Joseph A., Phillip J. Mantgem, Derek J. Young, Kristen L. Shive, Haiganoush K. Preisler, Adrian J. Das, Nathan L. Stephenson, et al. “Effects of Postfire Climate and Seed Availability on Postfire Conifer Regeneration.” *Ecological Applications* 31, no. 3 (2021). <https://doi.org/10.1002/eap.2280>.

Tattoni, Clara, Francesco Chianucci, Marco Ciolli, Carlotta Ferrara, Luca Marchino, Michele Zanni, Paolo Zatelli, and Andrea Cutini. “A Comparison of Ground-Based Count Methods for Quantifying Seed Production in Temperate Broadleaved Tree Species.” *Annals of Forest Science* 78, no. 1 (2021). <https://doi.org/10.1007/s13595-020-01018-z>.

THOMAS, JACK WARD, JERRY F. FRANKLIN, JOHN GORDON, and K. NORMAN JOHNSON. “The Northwest Forest Plan: Origins, Components, Implementation Experience, and Suggestions for Change.” *Conservation Biology*20, no. 2 (March 27, 2006). <https://doi.org/10.1111/j.1523-1739.2006.00385.x>.

Wright, Micah C., Phillip van Mantgem, Nathan L. Stephenson, Adrian J. Das, and Jon E. Keeley. “Seed Production Patterns of Surviving Sierra Nevada Conifers Show Minimal Change Following Drought.” *Forest Ecology and Management* 480 (2021): 118598. <https://doi.org/10.1016/j.foreco.2020.118598>.