

# Calculating Timber Harvest Costs Based Solely on Spatial Predictors Exemplified by the Colorado State Forest

Ulrich Strötz

ustroetz@gmail.com

Full paper double blind review

## Abstract

Optimization models for ecological forestry approaches require consideration of a variety of spatial features, including Harvest Costs, in order to maximize triple bottom line returns. Since the composition and the structure of the forest systems are usually not available for an entire landscape, a model is required that calculates Harvest Costs solely based on Spatial Predictors, which are Slope and Skidding Distance. Currently, no existing study investigates the significance of Spatial Predictors on Timber Harvest Costs. Therefore it is also not known if the significance of Spatial Predictors on Harvest Costs is high enough to calculate Timber Harvest Costs solely based on Spatial Predictors.

A dataset containing 160,000 test units based on existing harvest data of the Colorado State Forest is created. The dataset contains the Spatial and Non-Spatial Predictors of Timber Harvest Costs for each unit. Each unit is run through a created Harvest Cost Model, which is based on existing literature and equations. The Harvest Cost Model returns a Cost per ton for each unit. The created data are used to develop a spatially explicit regression model that calculates Harvest Costs solely based on Spatial Predictors.

The created spatially explicit regression model has an R-squared value of 0.42. Therefore Spatial Predictors predict 42% of Timber Harvest Costs. Calculating Timber Harvest Costs with an accuracy of 42% is not enough to calculate absolute Harvest Costs solely based on Spatial Predictors. But for optimization models, relative Harvest Costs are sufficient, since relative Harvest Cost allows the comparison of Costs of different stands and scenarios. An accuracy of 42% is then enough to estimate relative Harvest Costs.

## 1 Introduction

The returns from forest management vary dramatically across geographic space. This is to some degree because of the variability in the composition and structure of the forest systems, but also because costs of logging operations vary. Ecological approaches to forestry can be competitive; however, such approaches require consideration of a variety of spatial features, making this type of management far more complex than standard industrial approaches (e.g. harvesting 40 acres in a 40 year rotation cycle). Tools exist that facilitate this understanding and help landowners maximize triple bottom line returns on ecological forestry approaches (ECOTRUST 2014, ZUURING et al. 1995). Models also exist for estimating

Timber Harvest Costs for specific locations (HARTSOUGH et al. 2001, LOEFFLER et al. 2006). However, when planning operations across even relatively small size woodlots (<100 acres), the optimization of returns requires pre-existing knowledge of the potential costs for the entire landscape. Optimization models require iterating through millions of potential solutions and comparing results in terms of an objective function. Harvest Cost is a crucial factor in optimization models. But calculating Harvest Costs during the iteration process would slow down the optimization significantly. Therefore, a pre-generated dataset with the potential Harvest Costs for the entire landscape is required.

## Research Problem

Based on the stated existing models for estimating Timber Harvest Costs, research has been conducted to investigate what predictors influence Harvest Costs and what their significance in predicting cost is:

SILVERSTEIN et al. (2006) researched the influence of tree variables (volume and number of trees) on Timber Harvest Costs, but assumed fixed values for Slope and Skidding Distance. ARRIAGADA et al. (2008) researched the influence of Slope, Volume per Tree, and Trees per Acre on Harvest Costs, but ignored Skidding Distance as a variable by keeping it constant. KEEGAN et al. (2002) and LOEFFLER et al. (2006) researched the influence of Volume per Tree, Trees per Acre, and Skidding Distance on Timber Harvest Costs, but kept Slope constant.

No currently existing study, including the above, investigates the influence of all predictors on Timber Harvest Costs, including all Spatial Predictors. Spatial Predictors, which are Slope and Skidding Distance, are predictors that can be determined through Geographic Information Systems. Non-spatial Predictors, which are Volume per Tree and Trees per Acre, require fieldwork or other methods of determining their value.

Since no research has been conducted on the influence of all predictors on Timber Harvest Costs, the significance of Spatial Predictors on Timber Harvest Costs is unknown. Therefore it is unknown if it is possible to calculate Harvest Costs solely based on Spatial Predictors. Two major research questions result:

1. What is the significance of Spatial Predictors on Timber Harvest Costs?
2. Is it possible to calculate Timber Harvest Costs solely based on Spatial Predictors?

The significance of Spatial Predictors on Harvest Costs has never been fully researched. Given that the significance of Spatial Predictors is high enough to exclude the Non-Spatial Predictors, it is possible to calculate Harvest Costs exclusively with Spatial Predictors. Determining Harvest Costs solely with Spatial Predictors would allow determining the Harvest Costs for an entire geographic region at once by using Geographic Information Systems. No knowledge of the composition and structure of the forest systems – and therefore no fieldwork – is necessary. Knowing the Harvest Costs for an entire landscape would allow including that knowledge in the above stated optimization models.

## Response to the Problem

This paper will determine the significance of Spatial Predictors on Timber Harvest Costs and will create a cost equation that enables the calculation of Timber Harvest Costs solely based on Spatial Predictors.

This will be accomplished by creating a dataset containing 160,000 test units. The dataset contains the Spatial and Non-Spatial Predictors of Timber Harvest Costs for each unit. Each unit is run through a created Harvest Cost Model, which is based on existing literature and equations (FIGHT et al. 2006). The Harvest Cost Model returns a Cost per ton for each unit. The spatial and non-spatial input data are then used as independent variables in a multiple linear regression model, with the resulting Harvest Cost from the model as the dependent variable. From the created regression model, a spatially explicit regression model is derived that calculates Harvest Costs based solely on Spatial Predictors. From the spatially explicit regression model, a dataset with the potential Harvest Costs for the entire landscape can be created.

To limit potential influences and variations on the results, e.g. through different terrain or species, an area of interest was determined. Because of the availability of data, the homogeneity of the forest, and a fixed area, the Colorado State Forest (CSF) was selected. The single used harvest method in the CSF is clearcutting, which is conducted with ground-based machines.

This paper mainly summarizes the findings of the master thesis “Calculating timber harvest costs based solely on spatial predictors exemplified by the Colorado State Forest” (STRÖTZ 2014).

## 2 Model Development

A harvest model was developed to estimate relative Harvest Costs per ton (\$/ton). The model is based on the Fuel Reduction Cost Simulator (FRCS) (FIGHT et al. 2006) software. Relevant formulas for a ground-based mechanized-felling of whole tree systems were taken from the FRCS, and were written up in a Python script. The script allows iterations over the model in order to analyze it. The model consists of three processing activities: felling, transportation to the landing, and processing at the landing.

Costs for loading the trees on to a log-truck, moving expenses for the machines, and management overhead costs are not included in the model. The model assumes, a hard wood fraction of 0%, a wood density of 39 lb./ft<sup>3</sup> (lodgepole pine; dominating harvest tree), an average log length measuring 32 ft., and costs for machine and labor reflect Western Colorado standards.

For each of the three activities (felling, transporting, and processing), a cost per cubic foot (\$/ft<sup>3</sup>) is calculated. The costs are dependent on the input variables of the model:

- **Slope (S)** represents the incline or steepness of the timber stand (in %).
- **Skidding Distance (SD)** is the distance from the location where the tree is felled, to the landing (in feet).

- **Trees per Acre (TPA)** are the number of trees removed from the stand. Since clear cutting is the dominating harvesting technique in the CSF, all standing trees are assumed to be cut.
- **Volume per Tree (VPT)** represents the average volume per tree of the stand's trees (in cubic feet).

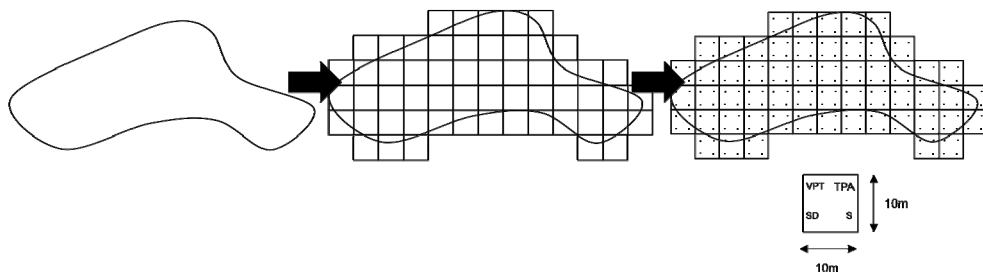
The calculated costs per cubic foot of each step are added together, resulting in a total Harvest Cost per cubic foot, and are converted to a Harvest Cost per ton.

### 3 Data Creation

For the statistical analysis, dependent and independent variables need to be defined and created. A dataset containing the independent variables, which are the input data to the cost model, and the dependent variable, which is the actual Harvest Cost calculated by the cost model, is created.

To make sure the created data are reasonable and represent realistic values for stands in the CSF, the input variables are derived from past timber sales of the Colorado State Forest Service (CSFS) (Townsend 2014). Data for 74 timber sales ranging from the years 2005 to 2014 are available from the CSFS including values for TPA and VPT.

Since Slope and Skidding Distance vary heavily across the stand, taking average values for the entire stand might bias the results. Therefore, the stands are split up into sub units of 10m by 10m. For each sub unit a value for each of the four variables will be produced and added to the dataset. Figure 1 illustrates the process on a sample stand.



**Fig. 1:** Data Creation exemplified on a sample stand

For each timber stand the average Volume per Tree and the average Trees per Acre value is available from the CSFS data. Since these values vary across the stand, a function, which draws random samples from a normal distribution around the average value of the stand, is run to produce a value for VPT and TPA for each sub unit that imitates the natural distribution across the stand.

From the National Elevation Dataset, which is available as a 10m raster from the U.S. Geological Survey (GESCH 2007), a slope raster in percent is derived. For each sub unit the zonal statistic of the Slope is calculated and added to the dataset.

Skidding Distance is the last value added to the dataset. The location of the landing is crucial to determine a reasonable distance. To imitate the common practice of creating several landings per stand, the stands are divided into four parts. For each part's centroid, the closest landing to an existing street in the OpenStreetMap database (OPENSTREETMAP WIKI CONTRIBUTORS 2014) is determined. In this way four landings, which serve the stand, are created. The Euclidian distance to each landing from each sub units' centroid is calculated, and the shortest distance is assigned as the Skidding Distance to the sub unit.

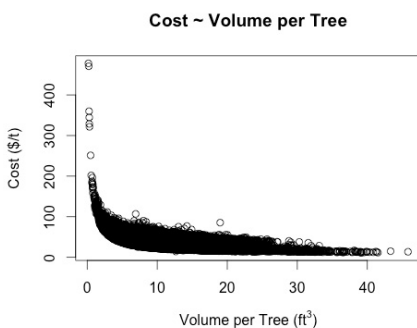
Overall 161,204 sub units were created with the described procedure covering the 74 stands. Since TPA and VPT are created with a function that draws random samples from a normal distribution, it is possible that the generated values are negative, though the values for TPA and VPT can by nature only be positive. Therefore, sub units containing negative numbers for TPA or VPT are removed from the dataset. Within the 161,204 sub units, this applied to 14 rows.

The last step in creating the dataset for the statistical analysis is to add the corresponding dependent variable to the independent variables. The input data are fed into the created cost model (Chapter 2), which returns a Harvest Cost per ton for each unit.

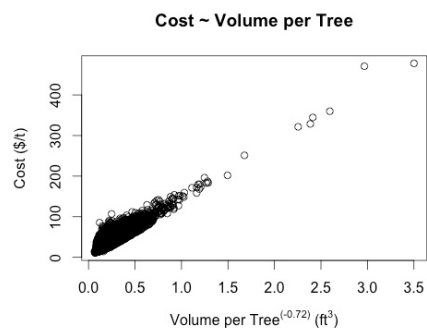
## 4 Statistical Analysis

The created dataset with the 160,000 sample units with values for each of the four input variables and the associated values for cost are statistically analyzed, and a regression model is built to model the relationship between the dependent variable cost and the independent input variables.

All independent variables, besides Volume per Tree have a linear relationship to cost. A linear relationship is necessary since the model is a linear regression model. A power transformation for Volume per Tree is used, in order to establish a linear relationship between cost and Volume per Tree. An exponent of -0.72 for Volume per Tree proved to be the best fit. Figure 2 shows the plot cost against Volume per Tree without the transformation, and Figure 3 shows the plot with the transformation of Volume per Tree.



**Fig. 2:** Plot Cost and Volume per Tree



**Fig. 3:** Plot Cost and transformed Volume per Tree

The following linear model, which respects all four independent variables as a linear relationship with cost, including the power transformation of VPT, is fitted in Equation 1.

$$C = -3.667098572 + 133.515209875 \times VPT^{(-0.72)} + -0.003088015 \times TPA + 0.305091203 \times S + 0.007587668 \times SD + \epsilon_i$$

**Equation 1:** Harvest Cost Regression with all predictors

Adding interaction terms to the model has been tested, but did not significantly increase the R-squared of the model.

From the developed multiple linear regression, a multiple-linear-spatially explicit regression model was derived using a backward selection. The non-spatial tree variables Trees per Acre and Volume per Tree were removed from the original model. With the ordinary least-square algorithm (OLS), the following model respecting only the two independent spatial variables as a linear relationship with cost is fitted in Equation 2.

$$C = 22.8038 + 0.3272 \times S + 0.007578 \times SD + \epsilon_i$$

**Equation 2:** Spatially explicit Harvest Cost Regression

#### 4.1 Model Validation

The F-statistics of both models have a p-value smaller than  $< 2.2e-16$ , which indicates that the models are likely to be significant. The p-values of the coefficients are below the significance level of  $\alpha = 0.05$  ( $< 2e-16$ ), therefore it is very likely that all the coefficients are significant.

The R-square value of 0.9828 of the model with all predictors indicates that the model is useful. The R-square value of the spatially explicit model is 0.4045, and also indicates that the model is useful. The model explains 40.45% of the variance of cost. The remaining 59.55% are not explained by the model, and will be discussed in Chapter 6.

The residuals of the models were explored in order to understand how well the model fits the data. Both models have a low magnitude of the median, indicating the majority of the residuals are around zero. The first quartile (1Q) and third quartile (3Q) have about the same magnitude in both models. The minimum and maximum values show outliers in both models, and will be discussed in Chapter 6.

Both models' data satisfies the assumptions of a linear regression. The residuals are plotted and show few high outliers, but also show that the majority of the values are located around the zero line, which suggests that the assumption that the data is linear is reasonable. Also, the residuals roughly form a "horizontal band" around the 0 line. This suggests that the variances of the error terms are equal (TEETOR & LOUKIDE 2011). The data satisfies the assumption of a linear model.

## 5 Results

A better understanding of the influence of the input variables on the Harvest Cost results from the statistical analysis. The generated input variables are all normally distributed, except for the variable Volume per Tree, as outlined in Chapter 4.

Overall, the four variables explain 98.28% of the Harvest Cost using the given model. All of them have an importance in predicting cost. The Trees per Acre variable, with 3.2% explanatory value, has the least importance, followed by Slope with 9.5%. Skidding Distance is the most important spatial variable with 29.9% explanatory value. And Volume per Tree is the most important variable out of the four, with 55.7% explanatory value. The two spatial variables taken together explain 39.4% of the cost.

The spatially explicit model was validated in Chapter 4. It is statistically significant and its coefficients are also statistically significant. The model is useful since it has a multiple R-squared value of 0.4212. To which degree it is useful, will be discussed in Chapter 6. The model shows outliers when the stands are extremely densely stocked (high Trees per Acre values), with a low timber volume (low Volume per Tree value). Also the model fits the data well and satisfies the assumptions of a linear regression.

The linear equation with the two variables and its coefficients is shown in Equation 3.

$$C = 22.83 + 0.3306 \times S + 0.007526 \times SD + \varepsilon_i$$

### Equation 3: Spatially explicit Harvest Cost Regression

The intercept of the spatially explicit regression is 22.83 \$/ton. This is the cost if the stand is in an absolute flat ground (Slope of 0 %), and the stand is located at the road (Skidding Distance of 0 ft.).

The coefficient for the variable Skidding Distance is 0.0076 \$/ft and the value for the Slope coefficient is 0.33 \$/%. Since both coefficients are positive, Harvest Cost will increase if Slope increases or the distance to the road increases. For each percent increase in Slope, the Harvest Cost will increase by 0.33 \$/ton. The Harvest Cost will also increase by 0.0076 \$/ton for each foot increase of the Skidding Distance. Or expressed in other units, the Harvest Cost will increase by 7.58 \$/ton for each additional 100 feet to skid.

## 6 Discussion and Conclusion

The research showed that Spatial Predictors predict 40% of Timber Harvest Costs. The remaining 60% are predicted by the variables Trees per Acre and Volume per Tree. Therefore the research question, which asks what the significance of Spatial Predictors on Timber Harvest Costs is, can be answered as follows: Spatial Predictors have a significance of 40% on Timber Harvest Costs. The gap in literature, which is caused by the lack of studies investigating the influence of all predictors on Timber Harvest Costs, including all Spatial Predictors, can be closed with the developed methodology and results of this paper.

The second research question, which asks if it is possible to calculate Timber Harvest Costs solely based on Spatial Predictors, depends on the use case: it is not possible to calculate an

absolute Harvest Cost with this method, because an R-squared of 0.4045 of the spatially explicit regression model is too low to calculate Harvest Costs solely based on Spatial Predictors. But this study was conducted in order to answer if it is possible to calculate Timber Harvest Costs for use in optimization models. Optimization models require iterating through millions of potential solutions and comparing results in terms of an objective function. For this kind of optimization, an R-squared value of 0.4045 is sufficient, because it indicates relative Harvest Costs. This allows optimization models to compare the costs of different stands and scenarios. These models do not require absolute Harvest Cost. Therefore, the results of this research make it possible to include Harvest Costs in optimization models for ecological forestry approaches. With their inclusion, optimization models are significantly improved.

The outlier pattern, that high Harvest Costs occur in extremely densely stocked stands with a low volume of the trees, is a realistic real world condition, and an expected silvicultural behavior. If stands are extremely densely stocked, the Volume per Tree value is low. Therefore the outliers were not removed from the dataset. But from a practical point of view, these stands are not likely to be harvested. It is likely that they caused part of the unexplained variance in cost. Future studies should take into account the exclusion of stands that show this outlier pattern.

The study also showed that Skidding Distance is a major factor in calculating Harvest Cost (29.9%). Skidding Distance heavily relies on the location of roads. As stated, roads usually are not created until a stand is harvested; which makes it very difficult to estimate the Skidding Distance properly. Therefore, further research should be conducted on ways to determine Skidding Distance precisely.

Also this research is restricted to a limited study area. The method of this study should be repeated in other areas and the results compared via further analysis.

## References

- ARRIAGADA, R. A., CUBBAGE, F. W., ABT, K. L. & HUGGETT, R. J. JR. (2008), Estimating harvest costs for fuel treatments in the west. *Forest Products Journal*, 58 (7/8), 24-30.
- ECOTRUST (2014), Forest Planner. <http://forestplanner.ecotrust.org/trees/about-the-forest-planner#about> (accessed 10 10, 2014).
- FIGHT, R. D., HARTSOUGH, B. R. & NOORDIJK, P. (2006), Users Guide for FRCS: Fuel Reduction Cost Simulator Software. General Technical Report. Pacific Northwest Research Station, Portland, OR.
- GESCH, D. B. (2007), The National Elevation Dataset. *Digital Elevation Model Technologies and Applications: The DEM Users Manual (American Society for Photogrammetry and Remote Sensing)*, 2, 99-118.
- HARTSOUGH, B. R., ZHANG, X. & FIGHT, R. D. (2001), Harvesting cost model for small trees in natural stands in the interior northwest. *Forest Products Journal*, 51 (4), 54-61.
- KEEGAN, C. E., III, NICCOLUCCI, M. J., FIEDLER, C. E., JONES, J. G. & REGEL, R. W. (2002), Harvest cost collection approaches and associated equations for restoration treatments on national forests. (Management). *Forest Products Journal*.



- LOEFFLER, D., CALKIN, D. E. & SILVERSTEIN, R. P. (2006), Estimating volumes and costs of forest biomass in western Montana using forest inventory and geospatial data. *Forest products journal*, 56 (6), 31-37.
- OPENSTREETMAP WIKI CONTRIBUTORS (2014). Overpass API. OpenStreetMap Wiki. [http://wiki.openstreetmap.org/w/index.php?title=Overpass\\_API&oldid=1085574](http://wiki.openstreetmap.org/w/index.php?title=Overpass_API&oldid=1085574) (accessed 10 10, 2014).
- SILVERSTEIN, R., LOEFFLER, D. P., JONES, J. G., CALKIN, D. E., ZUURING, H. R. & TWER, M. (2006), Biomass utilization modeling on the Bitterroot National Forest. *Fuels Management-How to Measure Success: Conference Proceedings*. Portland, OR: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 673-688.
- STRÖTZ, U. (2014), Calculating timber harvest costs based solely on spatial predictors exemplified by the Colorado State Forest. Master Thesis. Centre for GeoInformatics (Z\_GIS), Salzburg University, Austria.
- TEETOR, P. & LOUKIDE, M. K. (2011), *R cookbook*. O'Reilly Media, Sebastopol, CA.
- TOWNSEND, H. (2014), Colorado State Forest Timber Sales. Shapefile. Colorado State Forest Service, Walden, CO.
- ZUURING, H. R., WOOD W. L. & JONES, J. G. (1995), Overview of magis: a multi-resource analysis and geographic information system. fsrn-int-427. USForest Service, Ogden, UT.