Example lab report

Relationship of Remotely Sensed Data to Conifer Quadratic Mean Diameter

Susie Derkins

1 January 2024

Introduction

Remotely sensed data are relatively easy to obtain and can provide complete census information for certain variables over a large area. In 1998, a study was undertaken to examine the relationship between quadratic mean diameter in inches of conifers in a stand and the signal obtained from the fifth band in remotely sensed reflectance data, a band that represents reflectance of mid-infrared light[1].

Quadratic mean diameter (QMD) is used to derive wood volume. If band-5 reflectance could be used as a proxy for QMD, estimates of wood volume in a stand could then be made relatively rapidly and cheaply using remotely sensed data. Previous research with predictive models[2] indicates that the mean absolute error (MAE) should be less than 3 inches on a test set of data for band-5 reflectance to be considered a sufficient single linear predictor of quadratic mean diameter[3]. In addition, the USDA Forest Service has established that the R^2 for the linear relationship between two variables should be at least 0.75 for the model to be considered useful for prediction[4].

Methods

In 1998, the QMD of conifers was measured by field crews in 70 1-hectare plots as part of the Continuous Vegetation Survey (CVS). CVS plots are chosen for sampling each year based on a randomly located systematic grid of plots established by the USDA Forest Service in order to make inference to overall forest health and status in region 6 (Oregon and Washington). All 70 plots selected for the 1998 CVS sample in the Douglas-fir/Western Hemlock ecoregion of western Oregon were included in this study.

Also in 1998, remotely sensed reflectance values were measured over the entire region and recorded for each 25 x 25 m unit of land (pixel)[5]. The reflectance data for the 70 CVS plots were extracted for use in analysis. As the 70 plots are a random sample of all possible forested plots in the ecoregion, the scope of inference for this study is to all forested land in the Douglas-fir/Western Hemlock ecoregion of western Oregon in 1998[6].

The predictive performance of a linear model using band-5 reflectance to predict mean QMD of conifers was tested based on a random subset of held-out data. The test set was 5% of the original dataset. All effects of interest, including the linear slope, R^2, and MAE were estimated from that model. All analyses were done with R version 4.0.5 (2021).

[1] Clear statement of research question

[2] A predictive model is a model where the goal is predicting values of the response variable in other datasets. You may be used to seeing inferential models, where the goal is to estimate parameters of the statistical model (i.e., estimate and describe relationships).

[3] Statement of biologically important effect with justification

[4] Statement of biological / practical benchmarks for statistics calculated from the fitted model

[5] Description of how the study was conducted and how the data were collected

[6] Statment of scope of inference, justified based on sampling design and expert knowledge

The following statistical model was fit to the data:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where y_i is the observed QMD of conifers for the i^{th} training stand, $i = 1, 2, ..., 63$, x_i is the band-5 reflectance for the i^{th} training stand, and β_0 and β_1 are unknown parameters to be estimated. β_0 represents the expected QMD when band-5 reflectance is zero, while β_1 represents the linear change in the mean QMD for a one-unit increase in band-5 reflectance. It was assumed that $\epsilon_1, \epsilon_2, ..., \epsilon_n$ were independent and identically distributed normal errors with mean zero and variance σ^2. This model assumes a linear relationship between mean QMD of conifers and band-5 reflectance, and also assumes that the variance of QMD of conifers among stands with the same reflectance value is the same as the variance of QMD of conifers among stands with a different reflectance value[7].

Results

Assumptions were checked graphically using the residuals from the fitted model. The model assumption of a linear relationship between QMD and band-5 reflectance appeared to be adequately satisfied as was the model assumption of homogeneity of variance with respect to the predicted value (Figure 1)[8].

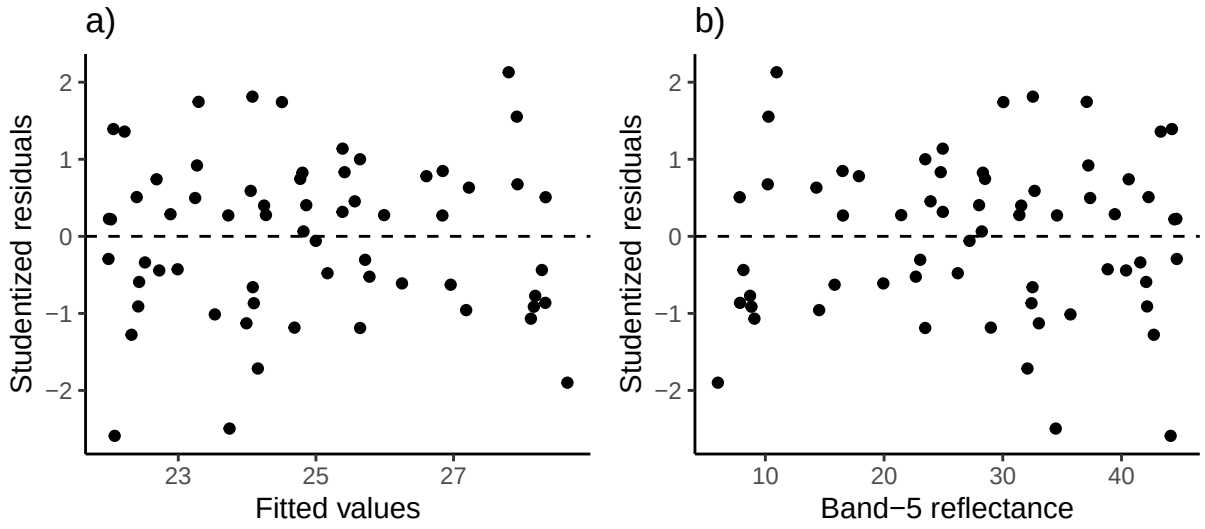

Figure 1: a) Studentized residuals versus the fitted values to check homogeneity of variance assumption. Because this assumption is reasonable met, prediction intervals are valid across the observed range of Band-5 reflectance. b) Studentized residuals versus the explanatory variable to check linearity assumption.

The fitted model predicts a decrease of 0.17 inches (95% CI [0.1 inches, 0.25 inches])[9] in the mean QMD[10] for every 1 unit increase in the band-5 reflectance (Figure 2). The adjusted R^2 for the fit to the training data was 0.25, while the MAE when predicting the test data was 5.7.

[7]Description of statistical model with all terms correctly defined

[8]Statement on how assumptions were checked and anything that was done if any assumptions were violated. Figures included in lab write-up as evidence assumptions were checked

[9]Confidence interval (with units) is valid because we only fit one model, selected *a priori*

[10]Notice that while the response is an individual value of QMD, (see model statement) the statistical interpretation is about the mean of the distribution of QMD.

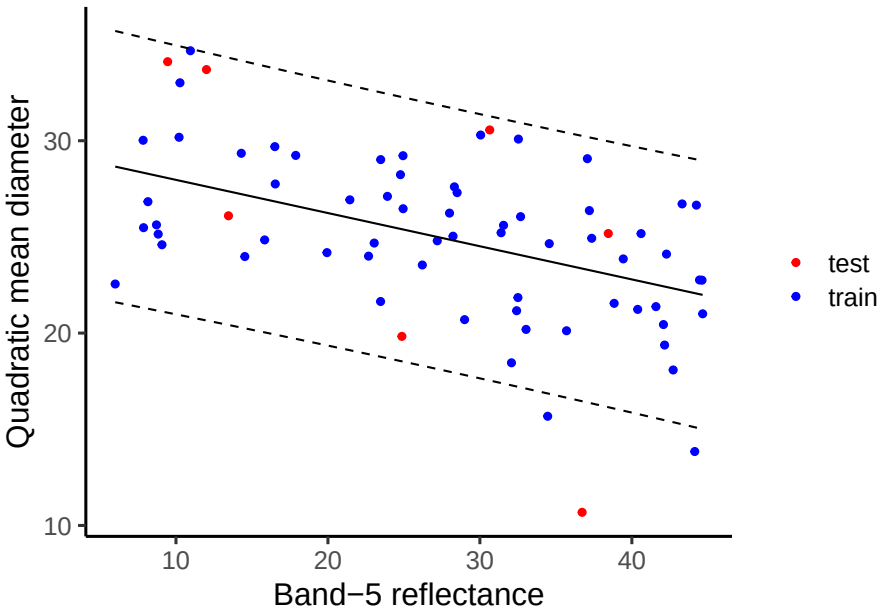

Figure 2: Fitted model with 95% prediction intervals. The observed data are overlayed and color coded to indicate which observations were used for training and which were used for testing

Conclusions

Although the relationship between mean QMD and band-5 reflectance was statistically detectable, the ability to predict an individual plot's QMD of conifers from band-5 reflectance alone was below the desired threshold considered useful (desired MAE was 3 inches, while MAE from the model was 5.7 inches). The adjusted R^2 for the fit to the training data was also below the desired threshold for the model to be considered useful (minimum R^2 for utility was 0.75, while the adjusted R^2 from the fit was 0.25)[11]. Predictive ability from this model gave 95% prediction intervals with a range of about 14 inches (Figure 2)[12]; this is too large to be of practical use for making estimates of stand wood volume. Based on these results, band-5 reflectance alone is not useful for making predictions of stand wood volume for stands in the Douglas-fir/Western Hemlock ecoregion of western Oregon in 1998, though it shows promise as a potentially useful predictor if in combination with other variables.

Further refinement of the model (e.g., inclusion of other remotely sensed explanatory variables) may improve the model's ability to predict QMD of conifers within established limits. In addition, further study on variability of remote sensing measures across years may be needed to make sure these results were not specific to the conditions in 1998. With data collected from additional years and environmental conditions, environmental variables could be collected and used in a more complex predictive model, as needed[13].

[11] Results and conclusions relate directly back to the purpose of the research

[12] A prediction interval is different from a confidence interval, and is appropriate when the goal is prediction. Make sure you understand why a prediction interval and not a confidence interval was used in the example.

[13] Limitation of project listed specifically. Further research possibilities discussed, with specific examples of what researches may want to focus on in the future.