

EDS 223: week 10 lab

Ruth Oliver

2023-12-05

Overview

The National Science Foundation's National Ecological Observatory Network (NEON) collects standardized, open-access ecological data at 81 freshwater and terrestrial field sites across the country. In addition to an amazing array of on-the-ground surveys, they also periodically collect Lidar data at the sites. All data is publicly available through the NEON Data Portal.

For this exercise, we will imagine that we are interested in studying canopy structure (tree height) at the San Joaquin Experimental Range in California. We're interested in figuring out if we can rely on the Lidar data NEON is collecting by comparing tree height estimates to on-the-ground field surveys. If the estimates between the two methods are similar, we could save ourselves a lot of time and effort measuring trees!

credit: this lab is based on materials developed by Edmund Hart, Leah Wasser, and Donal O'Leary for NEON.

Task

To estimate tree height from Lidar data, we will create a canopy height model (CHM) from Lidar-derived digital surface and terrain models. We will then extract tree height estimates within the locations of on-the-ground surveys and compare Lidar estimates to measured tree height in each plot.

Data

Lidar data

- Digital surface model (DSM) `SJER2013_DSM.tif`
- Digital terrain model (DTM) `SJER2013_DTM.tif`
- DSMs represent the elevation of the top of all objects
- DTMs represent the elevation of the ground (or terrain)

Vegetation plot geometries

- `SJERPlotCentroids_Buffer.shp`
- contains locations of vegetation surveys
- polygons representing 20m buffer around plot centroids

Vegetation surveys

- `D17_2013_vegStr.csv`

- measurements for individual trees in each plot
- metadata available in D17_2013_vegStr_metadata_desc.csv

Start by loading all necessary packages and setting working directory.

```
rm(list = ls())
library(terra)
library(sf)
library(tidyverse)
library(tmap)
library(here)

here::i_am("labs/week10-prep.R")
```

Load Lidar data.

```
# digital surface model (DSM)
dsm <- rast("./data/week10/SJER2013_DSM.tif")

# digital terrain model (DTM)
dtm <- rast("./data/week10/SJER2013_DTM.tif")
```

Check if the DSM and DTM have the same resolution, position, and extent by creating a raster stack.

```
test_raster <- c(dsm, dtm)
```

Create the canopy height model (CHM) or the height of all objects by finding the difference between the DSM and DTM.

```
chm <- dsm - dtm
```

Load the vegetation survey data, including the locations of study plots and the surveys of individual trees in each plot.

```
# read in plot centroids
plot_centroids <- st_read("./data/week10/PlotCentroids/SJERPlotCentroids_Buffer.shp")

# test if the plot CRS matches the Lidar CRS
st_crs(plot_centroids) == st_crs(chm)

tm_shape(chm) +
  tm_raster() +
  tm_shape(plot_centroids) +
  tm_polygons()
```

Load in the vegetation survey data and find the maximum tree height in each plot.

```
# read in the vegetation surveys, which include the height of each tree

# setting this option will keep all character strings as characters
options(stringsAsFactors=FALSE)
```

```
# read in survey data and find the maximum tree height in each plot
veg_surveys <- read.csv("./data/week10/VegetationData/D17_2013_vegStr.csv") %>%
  group_by(plotid) %>%
  summarise("survey_height" = max(stemheight, na.rm = TRUE))
```

Now find the maximum tree height in each plot as determined by the CHM.

```
extract_chm_height <- terra::extract(chm, plot_centroids, fun = max) %>%
  rename(chm_height = SJER2013_DSM) %>%
  select(chm_height)
```

Combine tree height estimates from the Lidar and plot surveys.

```
plot_centroids <- cbind(plot_centroids, extract_chm_height) %>%
  left_join(.,veg_surveys, by = c("Plot_ID" = "plotid"))
```

Plot results! Let's compare the estimates between the two methods (Lidar and on-the-ground surveys). To make the comparison, we'll add a 1:1 line. If all the points fall along this line it means that both methods give the same answer. We'll also add a regression line with confidence intervals to compare how the overall fit between methods compares to the 1:1 line.

```
ggplot(plot_centroids, aes(y=chm_height, x= survey_height)) +
  geom_abline(slope=1, intercept=0, alpha=.5, lty=2) + #plotting our "1:1" line
  geom_point() +
  geom_smooth(method = lm) + # add regression line and confidence interval
  ggtitle("Validating Lidar measurements") +
  xlab("Maximum Measured Height (m)") +
  ylab("Maximum Lidar Height (m)")
```

We've now compared Lidar estimates of tree height to on-the-ground measurements! It looks like the Lidar estimates tend to underestimate tree height for shorter trees and overestimates tree height for taller trees. Or maybe human observers underestimate the height of tall trees because they're challenging to measure? Or maybe the digital terrain model misjudged the elevation of the ground? There could be many reasons that the answers don't line up! It's then up to the researcher to figure out if the mismatch is important for their problem.