Colin White
March 17, 2021

HOMEWORK ANALYSIS #7

Models for snow density as a function of depth are essential to accurately track snow water accumulation over the recent past and to changes total ice sheet mass balance (whether ice sheets are gaining or losing water). Altogether, understanding snow density processes in Antarctica is essential for evaluating the health of the earth's largest fresh water reservoir. Your goal is to effectively model snow density cores and understand what types of patterns are present in the data. Ideally, this model would be used for both prediction and inference. The most important question you will address is whether each core needs its own mean function or whether the one curve is sufficient. For these data, you will be asked to consider various models and determine which is best.

The dataset "snow core.csv" gives the snow density for 14 snow cores. The variables in this dataset are described in Table 1. Each measurement represents the average density for a 2 cm section of the core.

Variable Name	Description
Core	The core or site index from which measurements are taken
Depth	Depth below the surface in m
Density	Density of snow in g/cm^3

Table 1: Variable descriptions

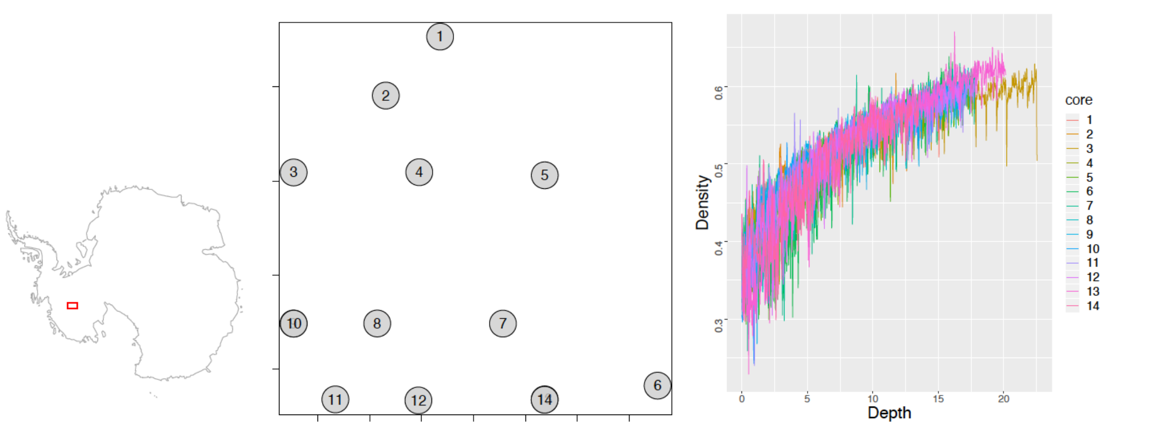


Figure 1: (Left) Region of Antarctica where the cores were drilled. (Center) Location within Antarctic region where cores are located. Note that cores 9 and 10 are very close. (Right) Snow density measurements.

1. Describe the goals of this analysis.

We want to know how snow density changes as a function of how deep under the surface of an ice sheet the observation is taken. In short, we want to be able to make inferences about snow density based depth. We also want to know if these inferences are valid between sites.

2. What is your outcome? What is/are your explanatory variable(s)?

The outcome will be snow density. The explanatory variables will be depth or some manipulated version of depth; and core number.

3. Fit a linear model using depth as a covariate. Fit a model with and without an interaction with core (remember to treat core as a categorical/factor variable) and discuss which is better. Something like lm(Density _ factor(Core) * Depth) should do the trick

The AIC for a linear fit model without interaction is -50348.49, The AIC for a linear fit model with interaction is -51851.09. Because the AIC for the model with interaction is smaller, it is better.

Code used to obtain results:

```
linear_mod = lm(Density ~ Depth + Core, data = snow_dat)
linear_mod_int = lm(Density ~ Depth * Core, data = snow_dat)
AIC(linear_mod)
AIC(linear_mod_inter)
```

4. Fit a cubic polynomial regression model using depth as a covariate. Fit a model with and without an interaction with core (remember to treat core as a categorical/factor variable) and discuss which is better.

The AIC for a cubic polynomial linear fit model without interaction is -55594.03, The AIC for a cubic polynomial linear fit model with interaction is -57519.7. Because the AIC for the model with interaction is smaller, it is better.

Code used to obtain results:

```
poly_mod = lm(Density ~ poly(Depth, degree = 3) + Core, data = snow_dat)
poly_mod_inter = lm(Density ~ poly(Depth, degree = 3) * Core, data = snow_dat)
AIC(poly_mod)
AIC(poly_mod_inter)
```

5. Fit a cubic spline model with a knot at the median of depth. Fit a model with and without an interaction with core (remember to treat core as a categorical/factor variable) and discuss which is better.

The AIC for a cubic spline linear fit model without interaction is -55616.3, The AIC for a cubic spline linear fit model with interaction is -57771.92. Because the AIC for the model with interaction is smaller, it is better.

Code used to obtain results:
knot_location = median(snow_dat$Depth)
spline_mod = lm(Density ~ bs(Depth, degree=3, knots = knot_location) + Core, data = snow_dat)
spline_mod_inter = lm(Density ~ bs(Depth, degree=3, knots = knot_location) * Core, data = snow_dat)
AIC(spline_mod)
AIC(spline_mod_inter)

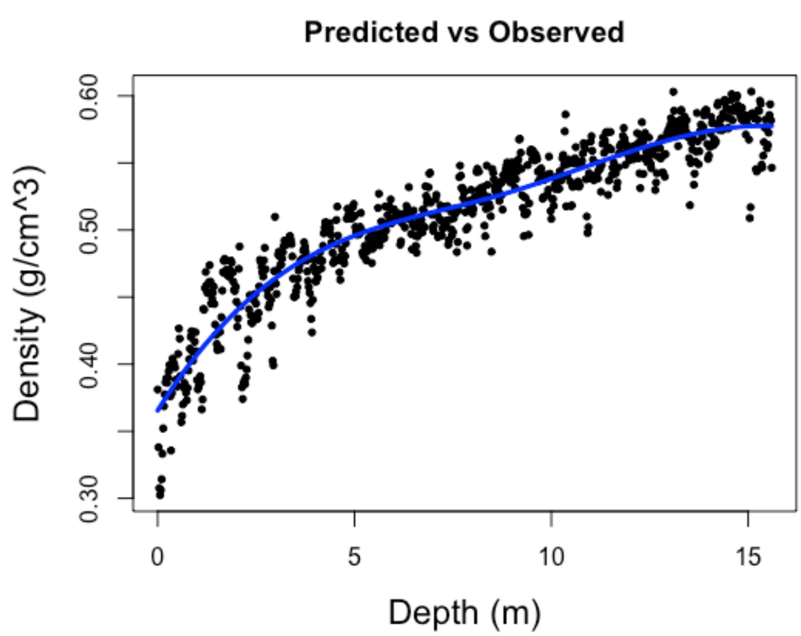
6. Of the six models fit, present your best model. Describe in words what your final model does. For example, is it one curve for all sites or a curve for each site? What are some advantages or disadvantages of your model?

The best model fit for this data is the cubic spline linear model with interactions. This is because it has the smallest AIC (-57771.92) of any of the 6 models. This model works by having a unique spline for each core.

7. For your final model, interpret Beta_0.

Beta_0 is the average expected estimated snow density for core being equal to 1, and the depth being equal to 0 (at the surface of the ice sheet).

8. For the first site (Core = 1), plot your model fit to verify that it is effectively capturing the snow density curve.



appendix:

```
#This code was derived from in class examples, and adapted using stack overflow resources.

library(ggplot2)
library(splines)
rm(list=ls())
setwd("~/Desktop/1A School/1A Winter 2021/STAT330/HW7")
snow_dat = read.csv("snow_core.csv")
snow_dat$Core = factor(snow_dat$Core)

#For problem 3
linear_mod = lm(Density ~ Depth + Core, data = snow_dat)
linear_mod_inter = lm(Density ~ Depth * Core, data = snow_dat)
AIC(linear_mod)
AIC(linear_mod_inter)

#For problem 4
poly_mod = lm(Density ~ poly(Depth, degree = 3) + Core, data = snow_dat)
poly_mod_inter = lm(Density ~ poly(Depth, degree = 3) * Core, data = snow_dat)
AIC(poly_mod)
AIC(poly_mod_inter)

#For Problem 5
knot_location = quantile(snow_dat$Depth,0.5)
knot_location = median(snow_dat$Depth)
spline_mod = lm(Density ~ bs(Depth, degree=3, knots = knot_location) + Core, data =
snow_dat)
spline_mod_inter = lm(Density ~ bs(Depth, degree=3, knots = knot_location) * Core, data =
snow_dat)
AIC(spline_mod)
AIC(spline_mod_inter)

# for problem 6
AIC(linear_mod)
AIC(linear_mod_inter)
AIC(poly_mod)
AIC(poly_mod_inter)
AIC(spline_mod)
AIC(spline_mod_inter)

#plot for problem 8
coef(spline_mod_inter)[1]
core1_fit = predict.lm(spline_mod_inter)[snow_dat$Core == 1]
plot(x = snow_dat$Depth[snow_dat$Core == 1], y = snow_dat$Density[snow_dat$Core==1],
    pch = 19, cex = 0.6, xlab = "Depth (m)", cex.lab = 1.4, main = "Predicted vs Observed",
```

```
     ylab = "Density (g/cm^3)")
lines(snow_dat$Depth[snow_dat$Core == 1], core1_fit, col = "blue", lwd = 3)
```