Predicting Tree Temperatures

Spencer Nelson, Danny Leybzon, Conor D'Arcy, Morgan de Ferrante

June 11, 2017

Contents

1. Abstract . . . 1
2. Introduction and Background of the study: Description of the study, and its background. 2
3. Questions to be Answered . . . 2
4. Description of the Sample and Data Collection . . . 2
5. Variables Studied . . . 2
6. Statistical Methods Used . . . 3
7. Summary of Findings . . . 4
8. Conclusion . . . 5
9. Shortcomings . . . 5
10. Recommendation for Future Research . . . 6

1. Abstract

Professor Sork and her team of dendrologists study acorn production from trees located at UCSB's Sedgwick Reserve. In the Sedgwick reserve, Professor Sork observes season changes of 103 different trees. On many of these trees, temperature sensors are located near the center of an 80m grid overlay. However, there are some that do not have these sensors. In order for Dr. Sork to better conduct her research, a full data set is desired. We were thus asked to help estimate temperature for trees that were missing buttons. We approached this problem by hypothesizing that we could use the trees with observed data to predict the data of trees without data. We made the general assumption that the closer two trees are, the more similar their temperatures would be. To further strengthen this assumption, we learned that a strong spatial pattern exists for variation in flowering and fruiting. Thus, our first model was created: Predicting temperature by proximity to other trees. However, we wanted to find the best predictions we could, so we attempted to build a second model. After comparing accuracy of the different models, we were able to reduce our error and accurately predict

temperature of trees. Since previous research has indicated that trees with early flowering are damaged by cold temperature, and trees with later flowering periods have pollen limitations, Professor Sork can now more accurately predict acorn yield trends.

2. Introduction and Background of the study: Description of the study, and its background.

UC Santa Barbara Sedgwick Natural Reserve has collected ten years of observations on springtime leaf and flower emergence of valley oak trees. Professor Sork has worked in conjunction with this department with the intention of understanding climate change's effects on the phenology of valley oaks. The study takes place in the Santa Ynez Valley of Santa Barbara County, and accounts for 103 adult valley oak trees located within about a 200 acre study area. Variables of interest recorded include leaf and flower observations (multiple observations per week from January 2007 to May 2015), acorn count, and temperature data as recorded by sensors attached (only about 50 trees had these sensors).

3. Questions to be Answered

1. How can we predict the temperature of trees using latitude and longitude accurately?
2. What is the best model for prediction?
3. How can we graphically provide the best insight?

4. Description of the Sample and Data Collection

The Sedgwick Reserve sample included 103 trees, with about 50 trees having temperature sensors. Leaf and flowering observations were made through on site visits around 1-2 times per week, and acorn production was counted using binoculars for 30 seconds. After data cleaning and removing observations with too many NA's for proper analysis, we were left with 36 trees with temperature sensors for our analysis, providing about 6 temperature observations per day from 2009 to 2016.

5. Variables Studied

% latex table generated in R 3.4.0 by xtable 1.8-2 package % Sun Jun 11 21:33:38 2017

	Variable	Description	Type
1	Year	Year of the observation	Categorical
2	Basedate	If we are analyszing for year 2009, then we choose Dec.1st.2008 as our base date	Categorical
3	Diff_1fstate1first	Number of days from previous year's Dec 1st to leaf bud burst time this year	Numerical
4	diff_catstat3first	Number of days from previous year's Dec 1st to catkin elongation time this year	Numerical
5	diff_flstate2first	Number of days from previous year's Dec 1st to flower formation time this year	Numerical
6	less35	The number of measurements below 35 degrees from previous year's Dec 1st to leaf bud burst time this year	Numerical
7	temp	The average temperature from previous year's Dec 1st to leaf bud burst time this year	Numerical
8	temp_split_low	Split daily temperature into high or low temperature group	Binary
9	Longitude	The longitudinal coordinate for a given tree	Numeric
10	Latitude	The lateral coordinate for a given tree	Numeric
11	Elevation	The elevation of a given tree	Numeric

6. Statistical Methods Used

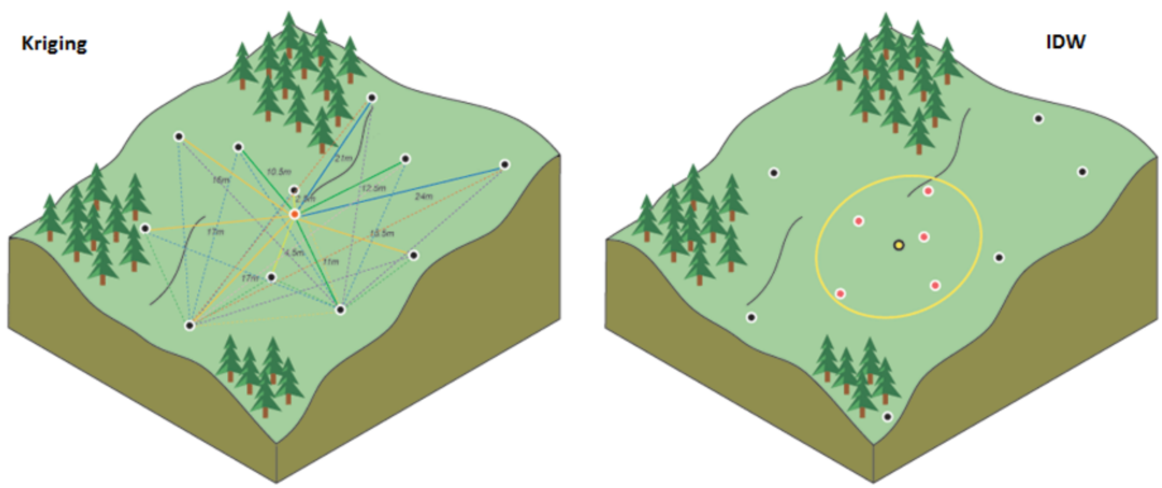

Figure 1: Kriging vs. IDW

Two Methods Compared:

We generated two models incorporating interpolation techniques called Kriging and Inverse Distance Weighting that could be used to generate temperature predictions of specific trees. Kriging and IDW weights the surrounding measured values to derive a prediction for an unmeasured location.

Kriging

The Kriging tool fits a mathematical function to data points within a specified area, to determine an output value for each location. Kriging uses a weighted average of neighboring samples, and these weights are optimized using the semi-variogram model, the location of the samples and all the relevant relationships between known and unknown values. Kriging provides the best linear unbiased prediction of intermediate values. It works by generating a predicted value at any unsampled location as a linear combination of the values and covariates observed at sampled locations.

$$\hat{Z}(x_0) = \begin{bmatrix} w_1 & w_2 & \cdots & w_N \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix} = \sum_{i=1}^{n} w_i(x_0) \times Z(x_i)$$

Figure 2: Kriging Formula

Inverse Distance Weighting

Inverse Distance Weighting is a type of deterministic method for multivariate interpolation with a known scattered set of points. The assigned values to unknown points are calculated with a weighted average of the values available at the known points by resorting to the inverse of the distance to each known point ("amount of proximity") when assigning weights. In other words, it gives greater weights to points closest to the prediction location, and the weights diminish as a function of distance. The method makes the assumption that things that are close to one another are more alike than those that are farther apart.

7. Summary of Findings

Summary of findings along with plots and summary tables along with the interpretation of results within the context of the study.

Tree Number	Latitude	Longitude	Temperature
19	34.70752	-120.04054	56.3
20	34.7073	-120.04025	61.40630861
21	34.70589	-120.04051	33.8
22	34.7056	-120.04067	57.1126333
24	34.7056	-120.04034	57.09989367
25	34.70545	-120.04027	57.40227772

Figure 3: Example Results

Figure 4 shows an example of predicted oak tree temperatures for a specified date and time. Our model can produce a temperature for any given date and time that is accounted for by the 36 trees in the dataset.

The map in Figure 2 plots trees over the designated area in the Santa Ynez Valley, with trees colored by

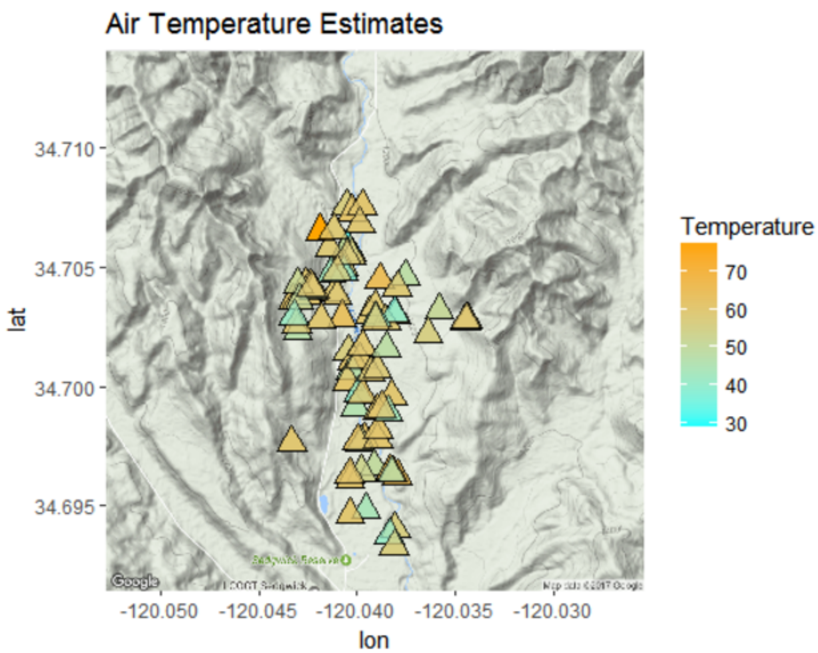

Figure 4: Predicted Temperatures

predicted temperature using the IDW method for a specified date and time. As you can see in the plot, temperature and patterns are correlated with elevation.

Figure 5 shows a snapshot of our interactive heat map (link provided below). The heat map shows predicted temperatures of oak trees for a specified date and time over the entire Santa Ynez Valley.

Heatmap: http://darcyconor.com/heatmap.html

8. Conclusion

In conclusion, we learned that Inverse Distance Weighting produces better estimates of temperature than Kriging does. Since our model relies on other trees to extrapolate temperature, we learned that despite having thousands of observations for individual trees, more trees would be necessary.

9. Shortcomings

While our model performed well, our mse of 18.96 was larger than we had hoped. Since our model relies on using other trees to predict the temperature of a given tree, we believe our error could have been reduced if we had access to more tree data. We also believe that our final model could have been further improved by

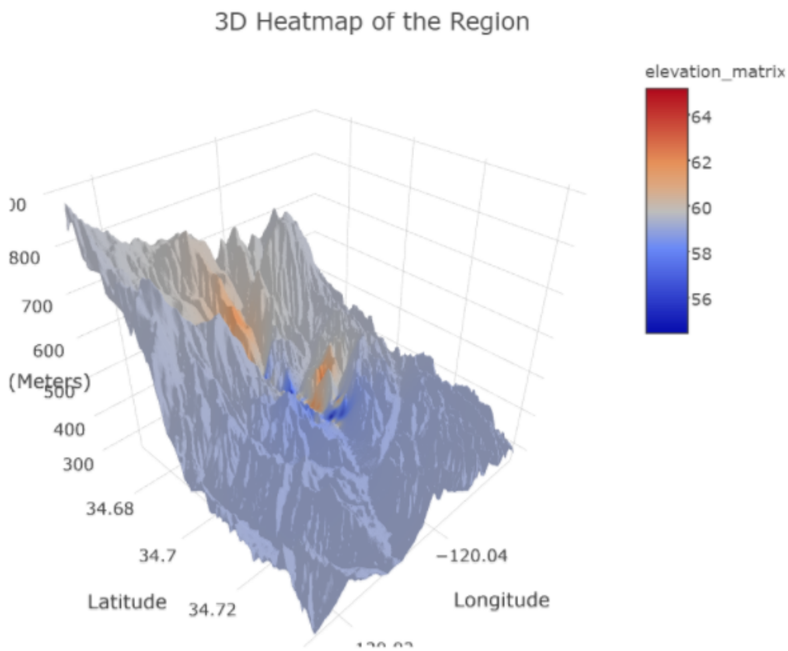

Figure 5: Heatmap

using external atmospheric data (pressure, water vapor, air composition).

10. Recommendation for Future Research

For future research we recommend that researchers collect atmospheric data as well as temperature, not only for better prediction of temperature but for better prediction of acorn production. If a researcher has access to these variables, we would recommend using a co-kriging model to include them in order to gain more accuracy.