

Report on Kriging in Interpolation

In this project I explored use of the geostatistical analyst extension and toolbar in the process of creating an interpolated surface through different kriging methods. In the tutorial, I focus on the parameters of the kriging surface and ability to improve interpolation through data analysis by comparing a 'trend removed' interpolated surface to a default kriging surface. Additionally, I compare the results of the kriging to other standard methods of interpolation like IDW, nearest neighbor, spline and trend. In this report I explore the use of interpolation in the context of ozone (O^3) concentrations in parts per million (ppm) throughout California. Ozone is used as a measure of air quality as in high concentrations it can cause adverse health effects.

The first part analysis I conducted involved the creation of an interpolated surface using default kriging parameters in the geostatistical analysis extension (Fig. 1). To create the default kriging surface I did not change the default parameters in the geostatistical wizard and selected ordinary kriging with no transformation or trend removal in a predictive output. The geostatistical wizard also outputs other auxiliary information which allows the user to explore trends and assumptions in the data. The first information output is a semivariogram which shows how related points are as a function of distance. The next output shows the search neighborhood for any point on the newly interpolated surface and shows the eight each data point has in interpolated point. The geostatistical wizard then outputs a cross validation which gives the user an idea of how well the model actually predicts the values at any location. The final output a numerical report of parameters and outputs from the tool. The interpolated surface then appears in the mapping area and can be used to extract ozone values at any location including cities.

In the next part of the analysis I used the geospatial analyst extension to explore the data in the ozone data collected at each point with the intention of identifying trends to better fit an interpolated surface. I first explored the distribution of data using the histogram tool to identify whether the distribution is normal or lopsided. Using the histogram tool, I identified the distribution of data as non-normal as the values are skewed right. Interpolated surfaces are most accurate over normally distributed datasets. I then used a quantile-quantile plot as another method of examining distribution of data. I observed in the distribution of ozone data that the primary deviations are in the lowest and highest values. To better identify this trend, I used the trend analysis tool in the geostatistical analysis. The trend analysis tool plots the spatial distribution with the magnitude of data and fits a mathematic function to the trend of data. Trends in the data can have many causes, which are not important to interpret yet, but it is clear in the ozone data that a second order polynomial function fits the east-west and north-south distributions of data. This observation is mirrored in the spatial autocorrelation and direction influence which is accessed through the semivariogram/covariance cloud tool in geostatistical analysis tool. Generally, the covariance cloud shows how similar points are as a function of distance. Using this diagram and the direction surface in the same tool, I observed that the points that are unusually different given their proximity are generally oriented east-west. In other words, the largest gradient in ozone concentrations is east to west, from the Central Valley to the San Francisco Bay area.

After exploring the data distribution I did another kriging interpolation using the geostatistical wizard (Fig. 2). Because the data had a second order trend from east to west and north to south, the order of trend removal was set to second. A directional influence is accounted for by adjusting the search neighborhood that is used to determine point values. The cross validation output from the tool shows qualitatively and quantitatively the increase in accuracy as the dataset now approaches a more normal distribution. By symbolizing the data sets by predicted standard error, it is easy to compare the amount of error in each interpolation (Fig. 3). I also compared the cross validation of each which shows the trend removed trend removed surface as more accurate (Fig. 4).

The next map I created is an indicator of probability that ozone concentrations crossed a threshold of 0.09 ppm which can be unhealthy (Fig. 5). I created this map by through kriging, but with an indicator kriging and probability output rather than ordinary kriging and predictive output as before. The final maps I created use other interpolation techniques to compare with kriging. One of the merits of the geostatistical analysis toolbar is the statistical data that assesses accuracy compared to a standard distribution. The other tools do not include this kind of option, so my comparison is restricted to objective observation. Inverse distance weighted (IDW) appears to have most favorable result of the other techniques (Fig. 6). Generally it shows the same trend with higher concentrations in the Central Valley and lower concentrations on the coast, but concentrations are less variable with a tighter distribution of values. The spline surface generally overestimates the values of ozone concentrations (Fig. 6). The trend surface does not appear to take into account the specific locations of values, and shows the overall trend of the data which increases in concentration to the north east (Fig. 6). The natural neighbor surface appear to not estimate outside the polygon area despite the processing extent set to the entire state, though the areas that are interpolated are fairly similar to the kriging surface, but shows lower concentrations (Fig. 6).

This type of analysis could potentially be useful for many studies over large areas where a continuous dataset needs to be extrapolated from discrete values. Air quality is a great example, but water quality and epidemiology are other fields where this type of study is useful. In my term project, interpolation is used to create an original hill surface on Cleman Mountain, from which incision depths are determined. The geostatistical analysis extension is particularly useful because it has so many parameter and additional statistics like error estimations.

Default Kriging of Ozone Concentrations in California

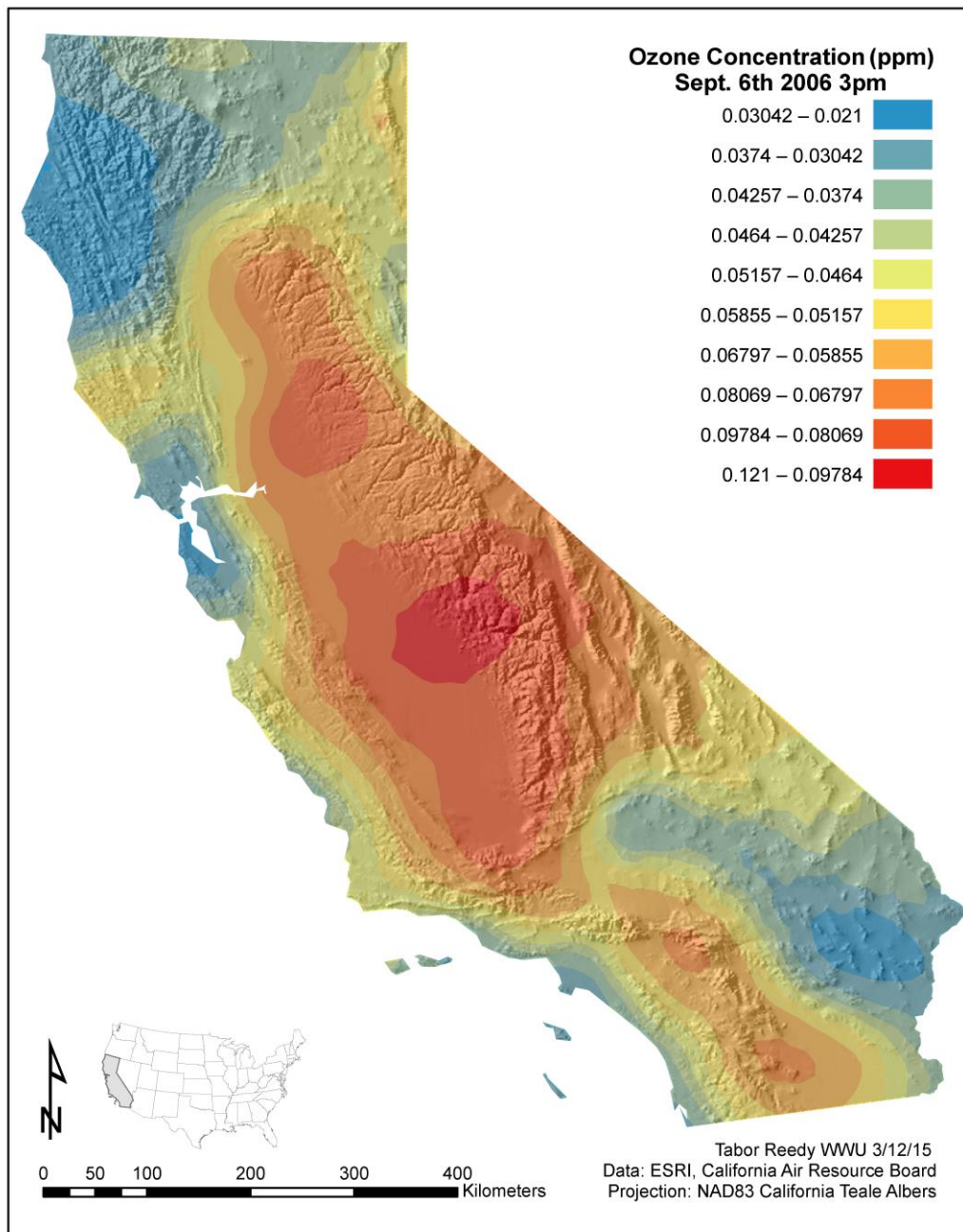


Figure 1. This map of ozone concentrations in California was created using default parameters in the geostatistical analyst extension in ArcGIS. Air testing locations were used with a kriging interpolation to create a continuous surface of predicted values. Kriging is a form of Gaussian regression governed by prior covariance. No trend is removed to account for directionality of the data distribution.

Trend Removed Kriging of Ozone Concentrations in California

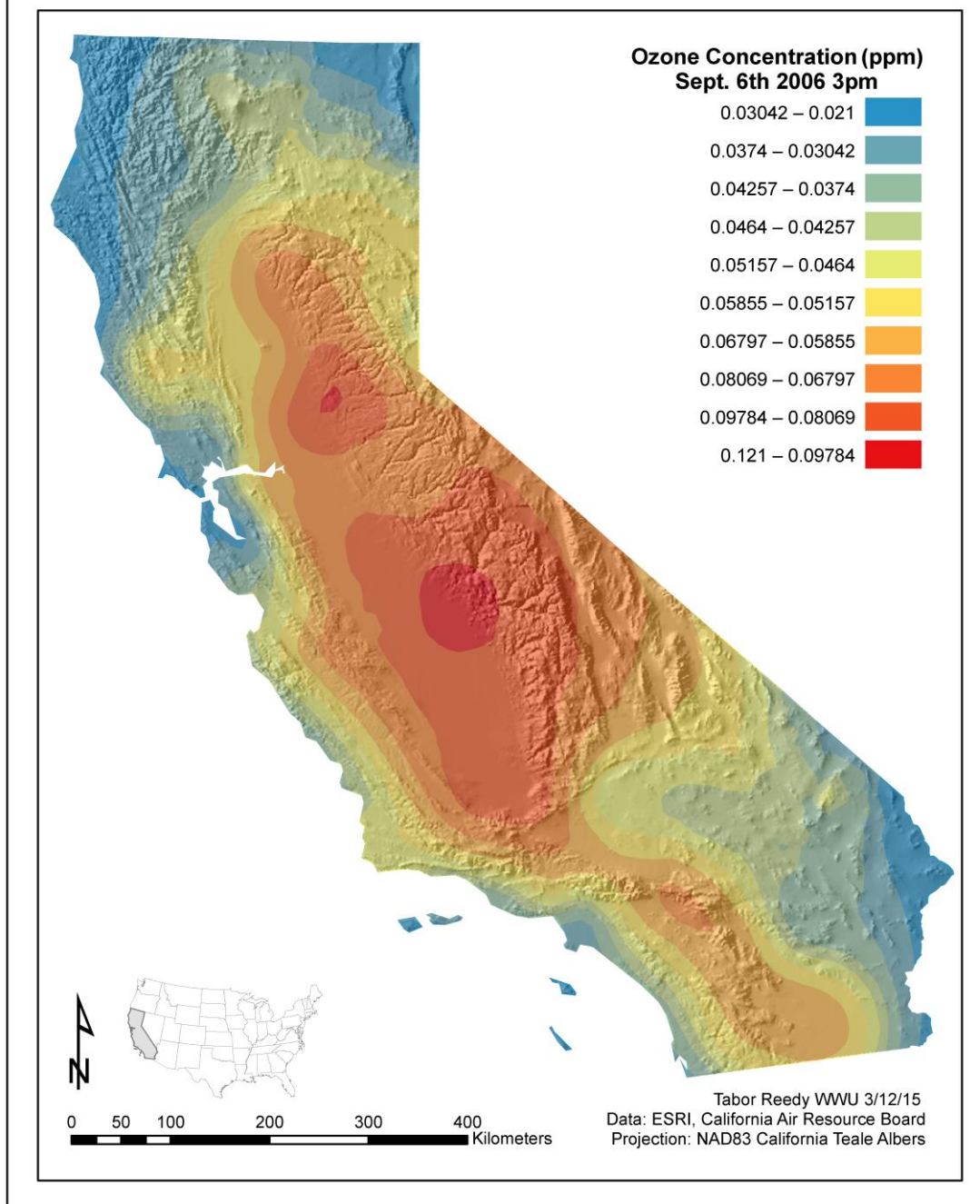


Figure 2. This map of ozone concentrations was created using a kriging interpolation with a second order polynomial trend removed and the neighborhood to weigh values adjusted for directionality of the data. This means the distribution generally accounts for high concentrations in the Central Valley, and low concentrations due to elevation in the mountains and wind on the coast.

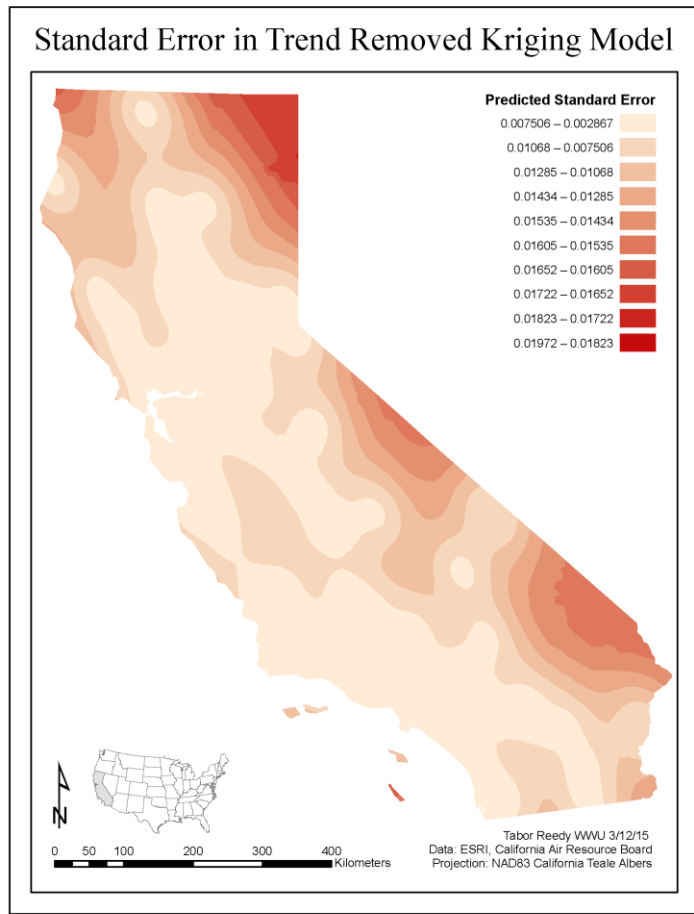
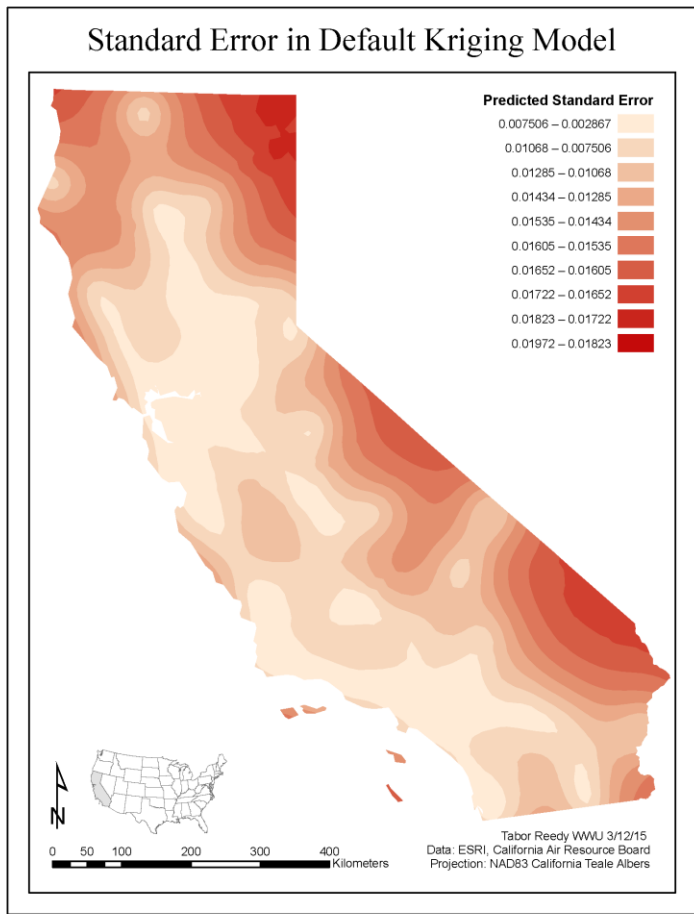


Figure 3. Objective standard error predictions based on distributions of data compared to a normal distribution of data. Interpolation techniques are best suited to normal distributions of data. By removing trends and adjusting for directionality, the trend removed surface is clearly more accurate.

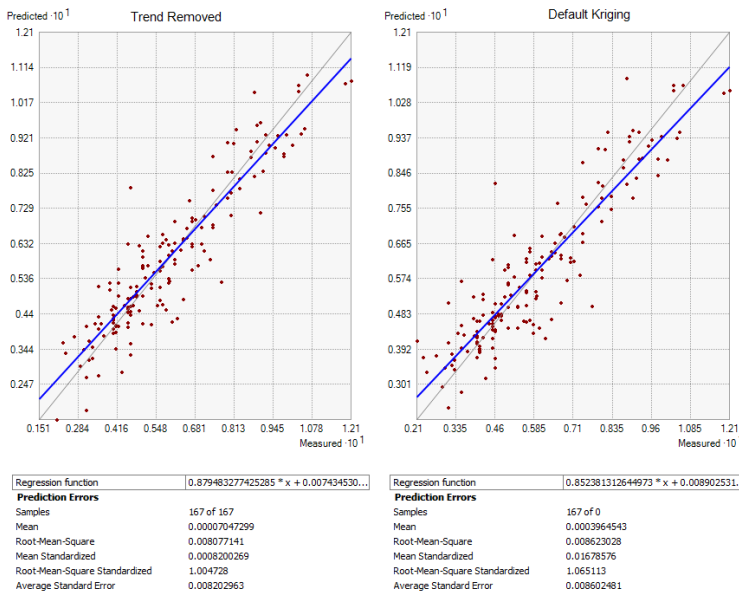


Figure 4. Distribution of data and the standard error associated with each interpolation. The trend removed kriging shows more accuracy, and the average line (blue) is closer to the linear normal distribution of data.

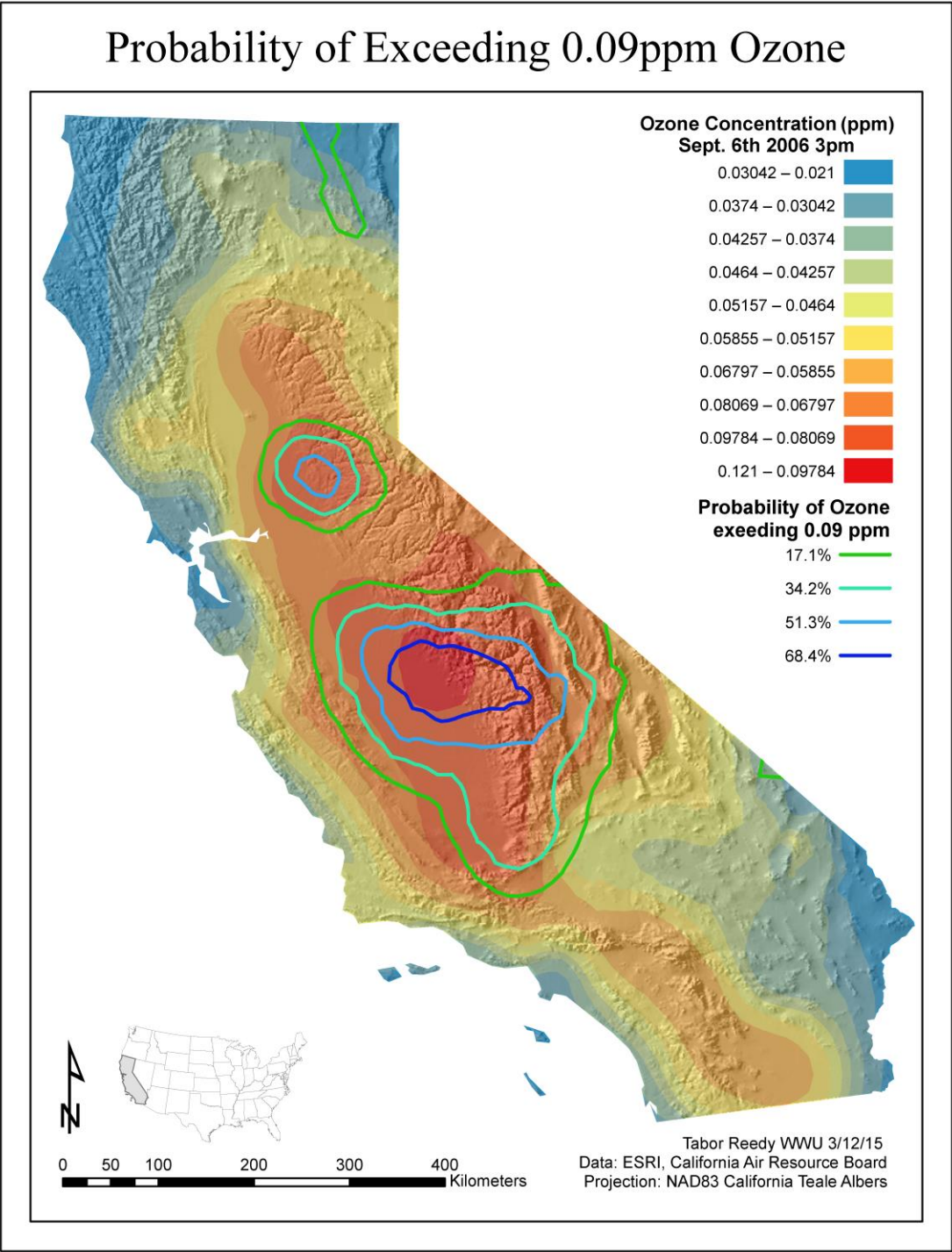


Figure 5. Probability that any area exceeded 0.09ppm over the 3-4pm time period on September 6th 2006. The interpolated surface is the trend removed kriging surface. Probability is calculated using an indicator kriging with a probability output. Kriging interpolation generated using the geostatistical analyst extension in ArcGIS.

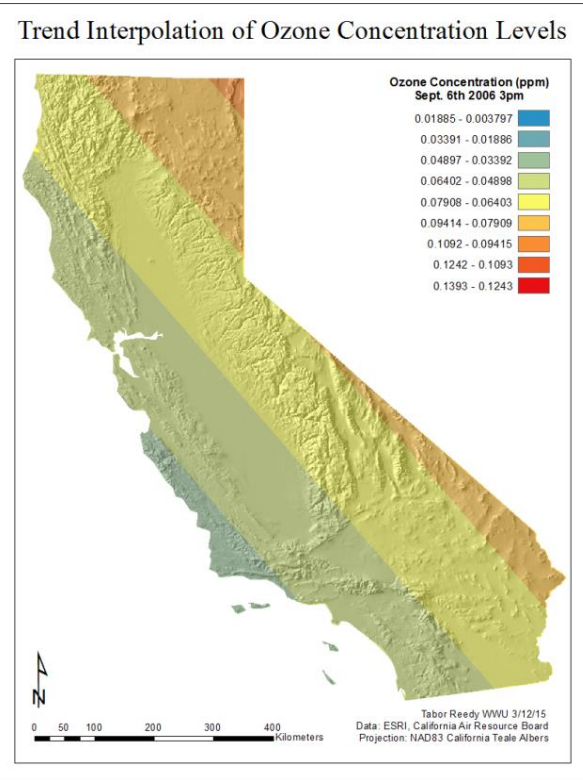
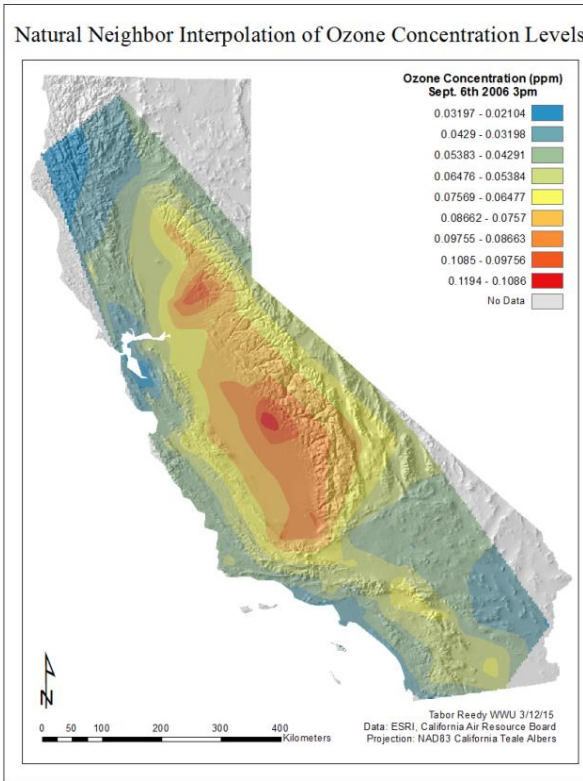
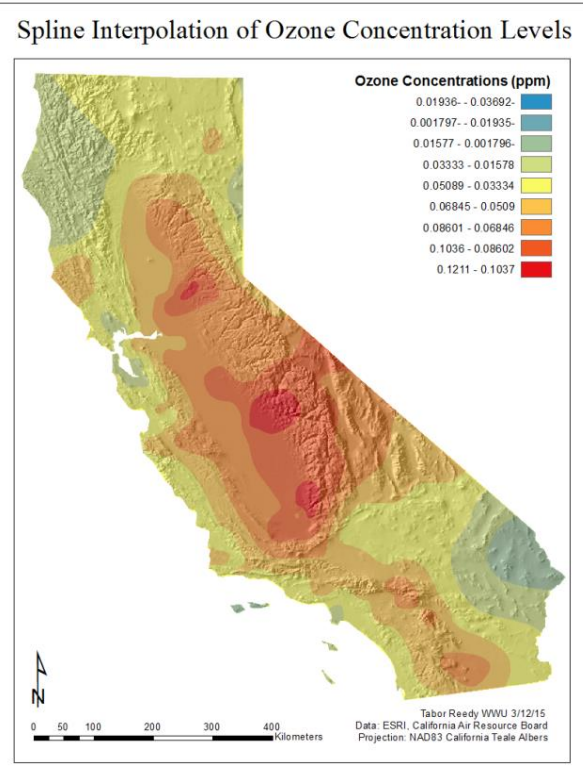
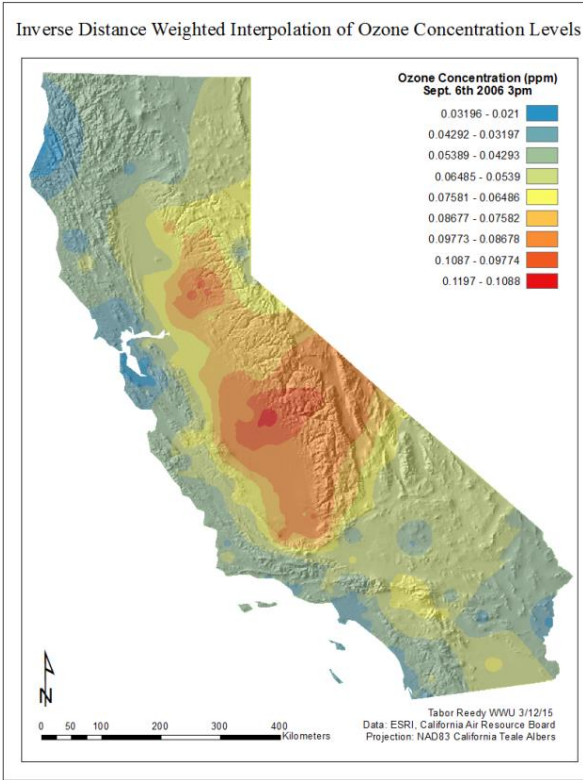


Figure 6. These are interpolated surfaces of ozone data using the same data set as the kriging interpolations. These are standard interpolations included in the spatial analyst extension in ArcGIS. IDW and spline interpolation seem to be the most accurate methods of classification especially when compared to the kriging interpolation. Natural neighbor and trend interpolations find their utility in smaller datasets with specific objectives such as knowing the overall trend of the data.