Albedo in Mexico

Sarah Jarvis
Spatial Statistics Homework 5

June 7, 2022

1 Exploratory Data Analysis

Below is a plot of all of the albedo measurements from the two satellites as well as the 1000 randomly selected points we used to fit the Gaussian process.

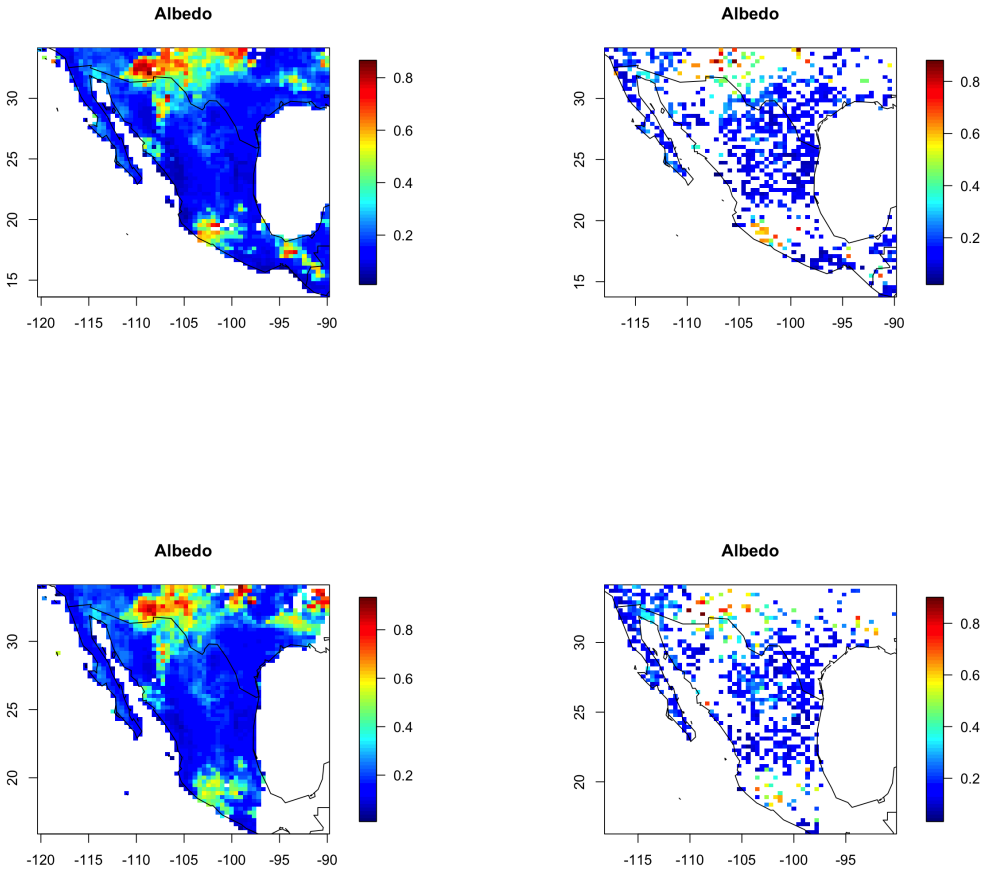
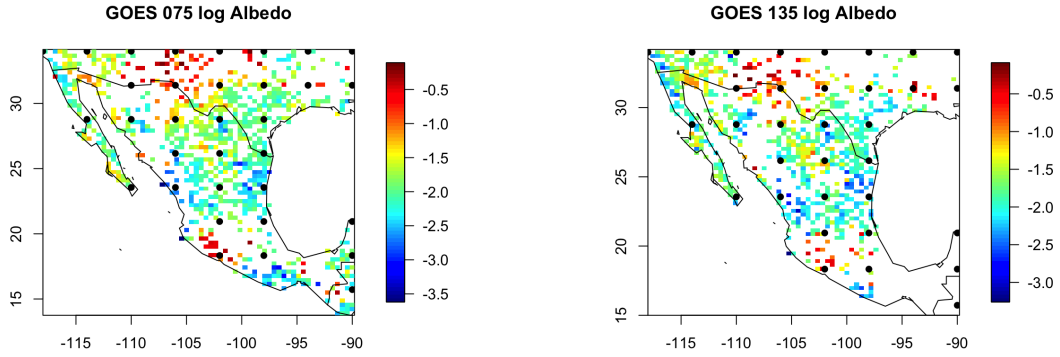


Figure 1. Observed albedo (left) and 1000 randomly selected albedo measurements from the same satellite (right). GOES 075 (top) and GOES 135 (bottom).

We can see some immediate differences between the observations. GOES 135 doesn't include the bottom right portion of Mexico and we see more red (high albedo) in the bottom of Mexico in GOES 075 than in GOES 135.

We use a log transformation of albedo because both box-cox and logit failed to normalize the variable. Below we can see the transformed albedo on the map as well as a black dot for the predictive process knots we will use for both satellites.



2 Trend/Model Details

We found evidence of a quadratic trend in the data for both satellites and selected the following model based on AIC with all of the covariates significant:

$$log(albedo) = \beta_0 + \beta_1 lon + \beta_2 lat + \beta_3 lon^2 + \beta_4 lat^2 + \beta_5 lon : lat$$

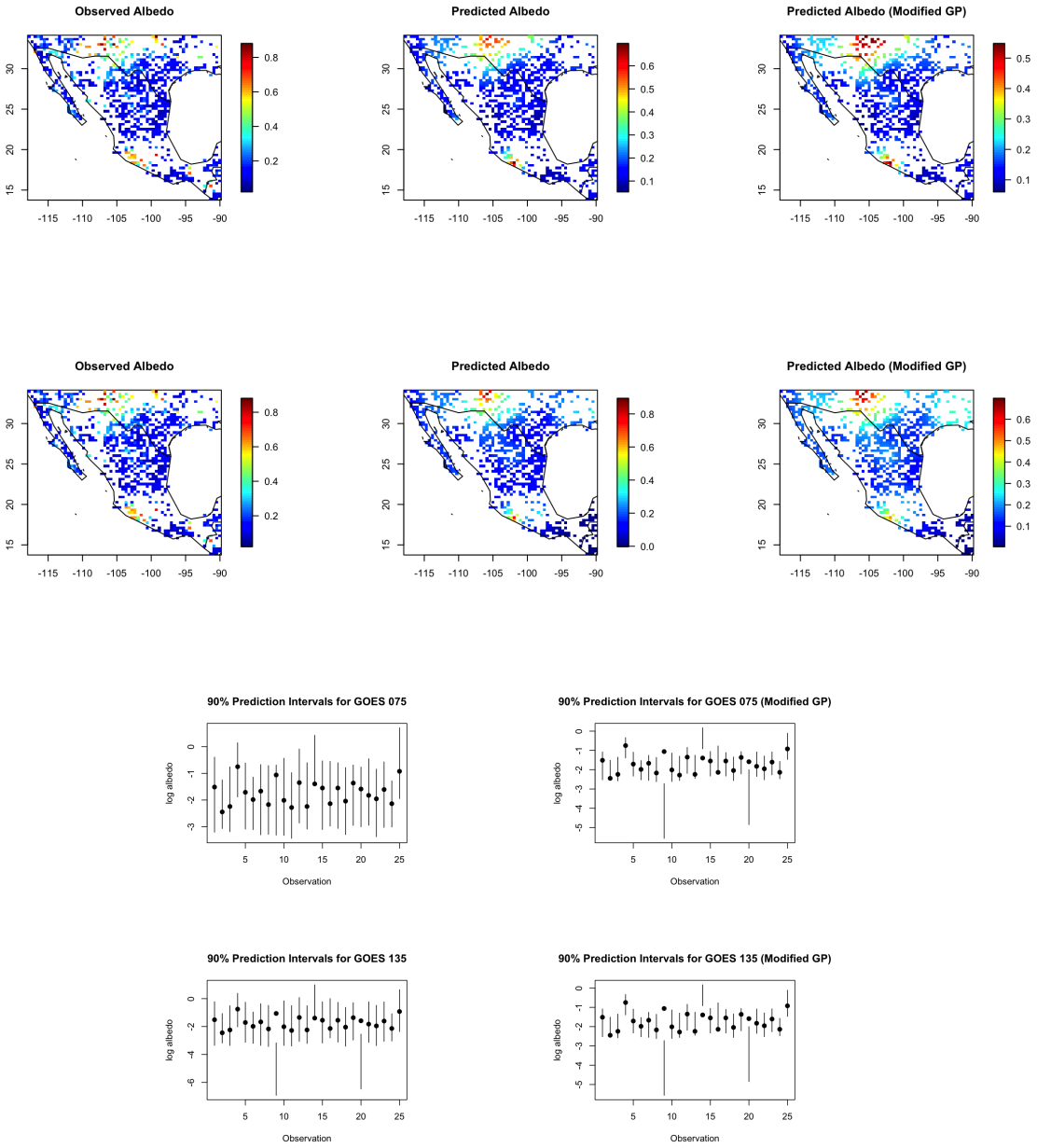
We used euclidean distance on latitude and longitude because our data is close enough to the equator for the distance to not be effected. We see this approach works with the accuracy of our predictions.

We select a smoothness parameter of 0.5 based on the value that gives the lowest AIC.

3 Model Check

The following methods were used to assess the models. The Gaussian process was fit using 1000 randomly selected points from the full data set. This is essentially our training set. A second set of 1000 randomly selected points was used as our testing set. We predicted on the

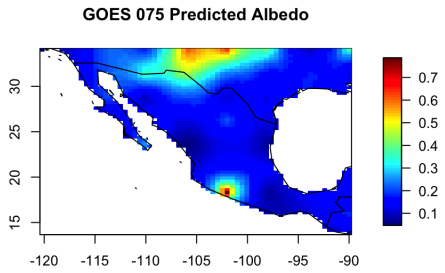
coordinates of the testing set and compared the output to what was observed thus being able to compare predictions to observations that were not used to fit the model. Below we can see that the predictions look very similar to the data. For GOES 075, we found that 99.2% of the testing observations fell inside the prediction intervals for the non-modified predictive process and 91.9% for the modified process. For GOES 135, we have a 94.5% accuracy for the non-modified and 85.6% for the modified process.

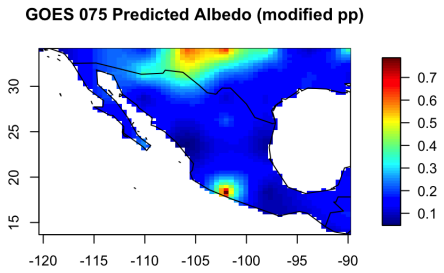


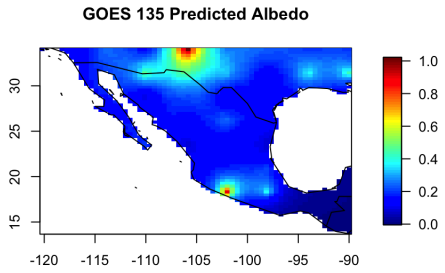
Testing set observations (left), predictions (middle), modified process predictions (right) for GOES 075 (top) and GOES 135 (bottom) and 90% prediction intervals with testing observations, non-modified (bottom left) and modified (bottom right).

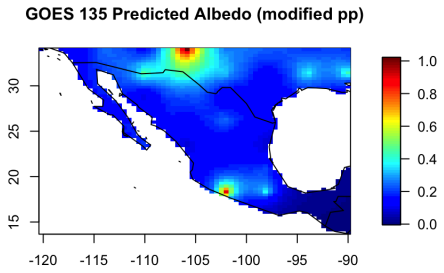
4 Predictions

Predicting over the whole area we obtain the following for each model and satellite. The predictions for the non-modified and modified look almost identical. The predictions look relatively similar to the original data. To get a better prediction we could use more knots or more randomly sampled data for fitting the model. Since we get a better accuracy with the non-modified, we will continue the analysis with the non-modified process.



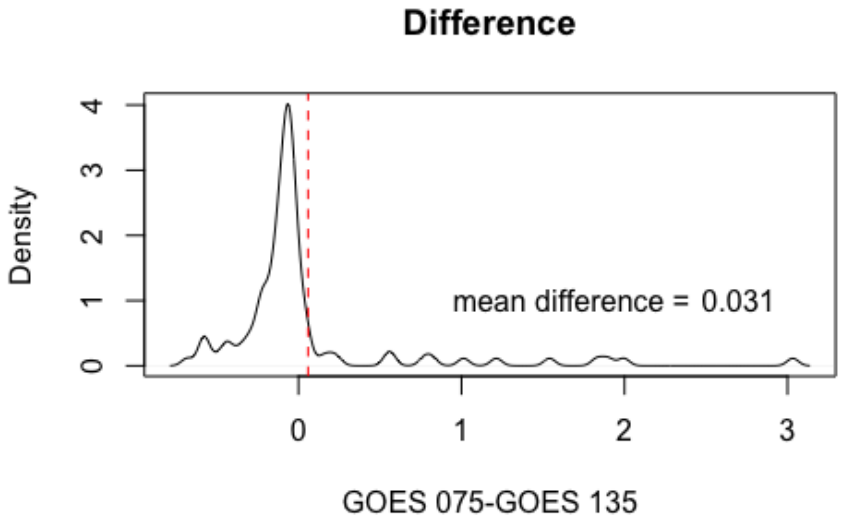


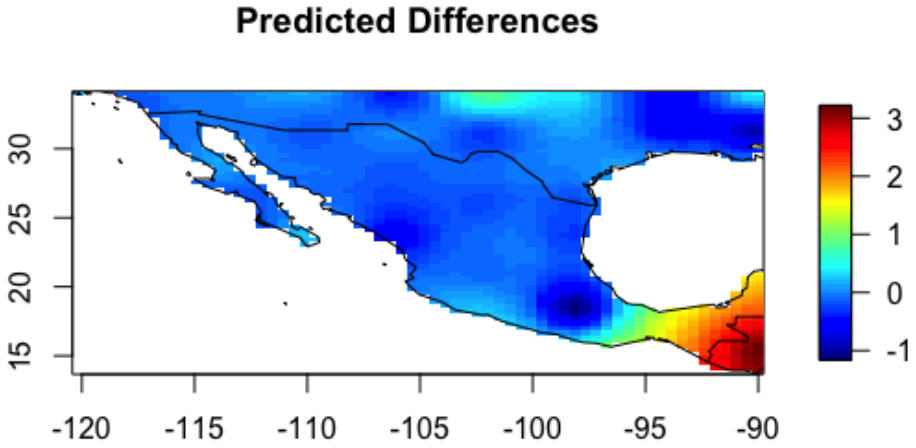




5 Satellite Differences

Randomly selecting 100 predictions and taking the difference results in the following distribution. We use the differences to fit a GP and obtain predicted differences for the whole grid. We can see that the major differences are in the bottom right corner. This is expected because we only have data in that area for one of the satellites.





Since the major differences are in the area that GOES 135 did not cover. I would continue any analysis with the GOES 075 data.