Albedo in North & Central America

Sarah Jarvis
Spatial Statistics Homework 4

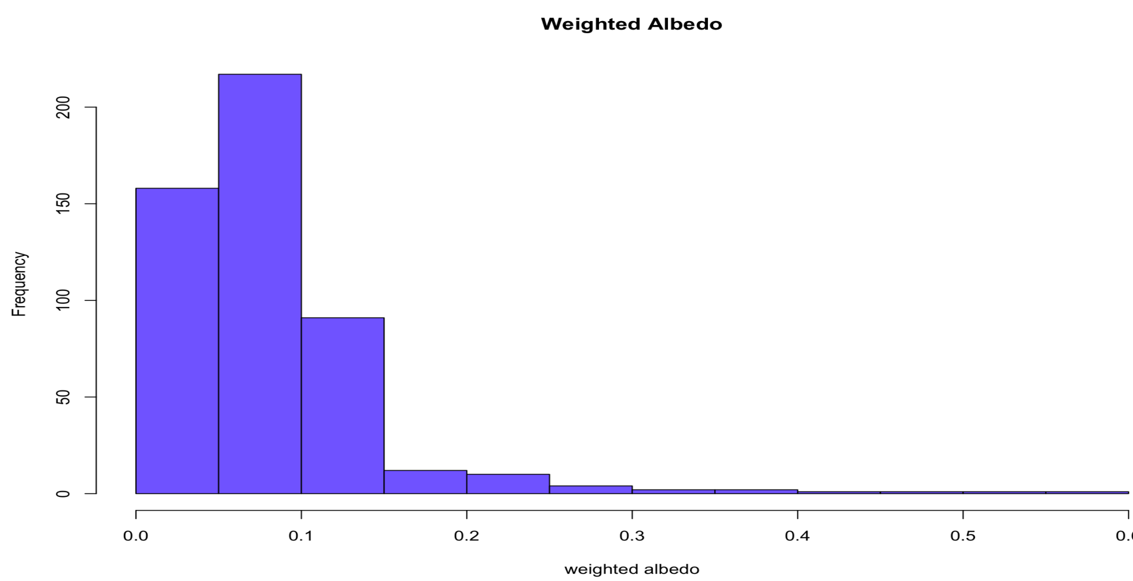
May 15, 2019

1 Abstract

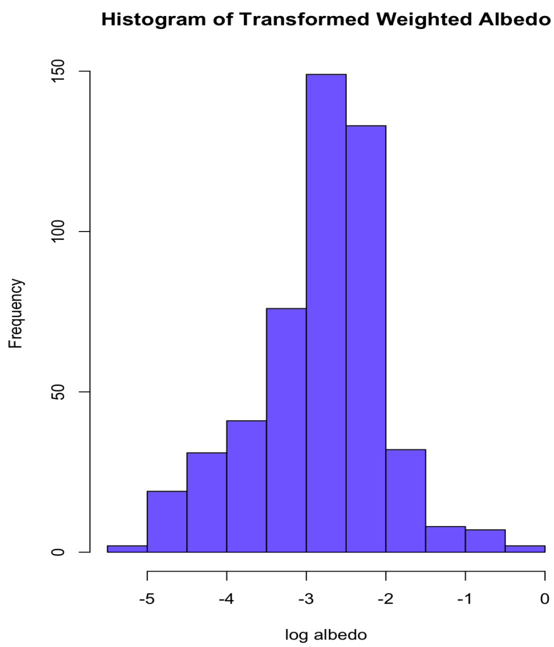
Albedo is a measure of how much light that hits a surface is reflected without being absorbed. It is important to study albedo as it has a significant contribution to global climate change. We fit a Gaussian process with a Matèrn correlation function to the albedo measurements at 500 locations in Central America and the eastern United States.

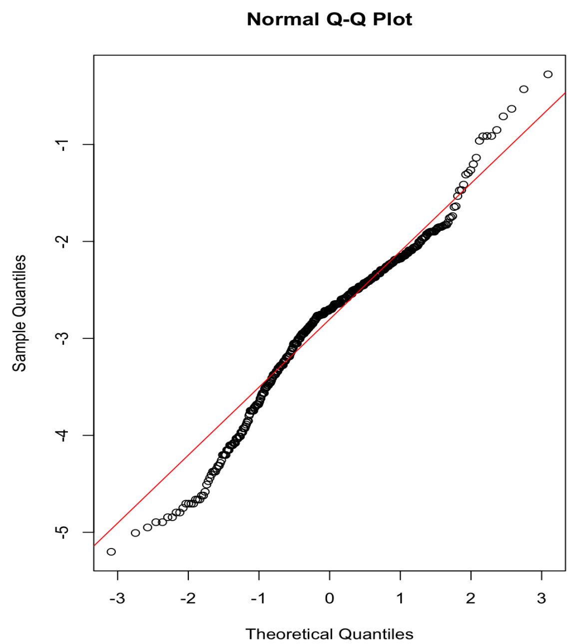
2 Data

The data for this analysis consists of 143,713 locations characterized by latitude and longitude as well as the albedo measurement and probability threshold corresponding to an index that indicates the quality of the retrieval of the information. We randomly sampled 500 of these locations for simplicity and a faster running analysis. We multiplied albedo by the probability threshold to get a weighted albedo measurement. Below we can see the distribution of this variable.

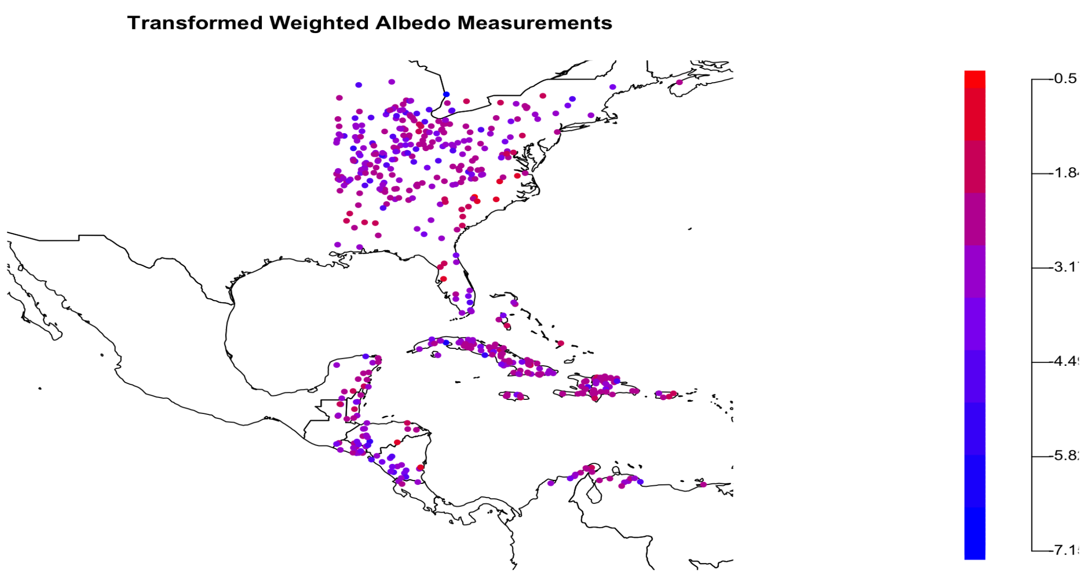


We can see that the distribution of weighted albedo is very right skewed. We take a log transformation in order to make the data closer to normality. We can see from the plots below that the transformation helped the distribution reach symmetry with slight difference from normality around the tails.





Below we show the transformed weighted albedo measurements on a map. Since the points were randomly selected, they should be a representative sample of the full data set for this region.



3 Methods

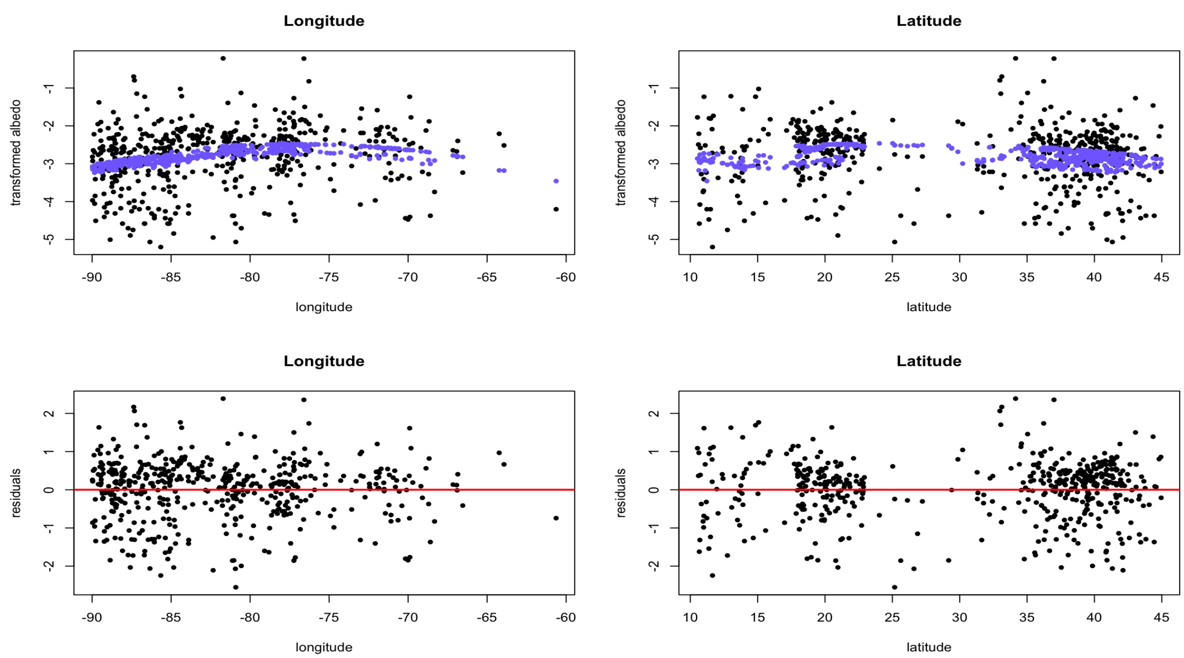
We explore possible trends in the data and fit a quadratic model selecting coefficients based on AIC. We then obtain the residuals and determine a fitting covariance function and explore possible anisotropies using a directional variogram. Finally, we fit our model to the data predicting albedo over the whole region and perform a goodness of fit by comparing our prediction intervals to the observations.

4 Trends

We plot log albedo against both longitude and latitude to explore the possible trends in the data. We fit the following model to the data and chose covariates based on AIC values. All of the following coefficients are significant in predicting albedo although the interaction between latitude and longitude was not. The data (black) and predicted (purple) values are shown below. We also show the residuals and see that the slight trend has been removed.

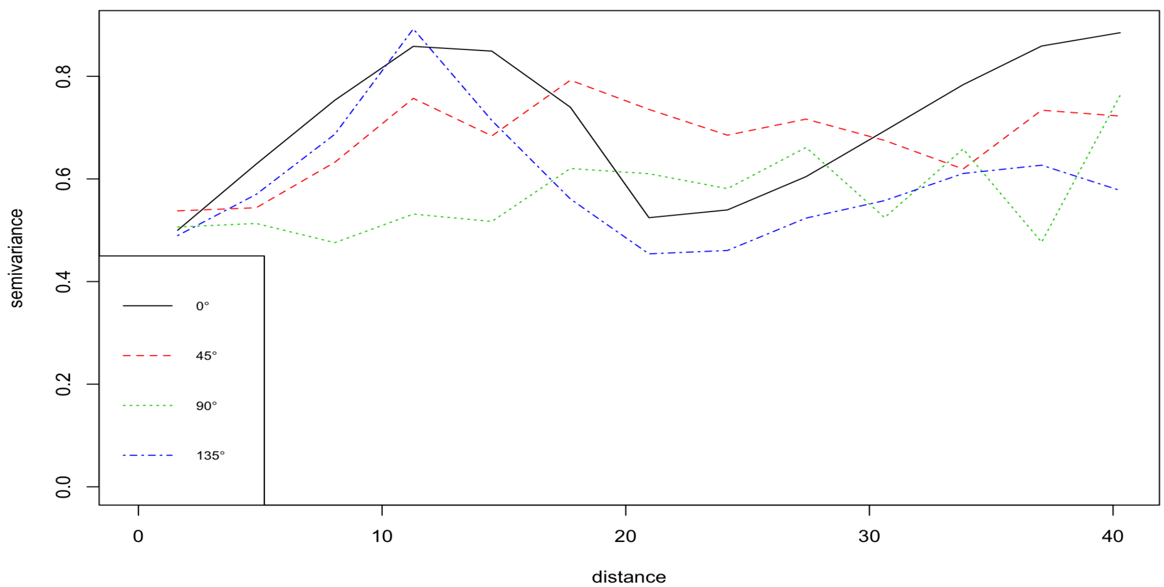
$$log(\hat{albedo}) = \beta_0 + longitude\beta_1 + latitude\beta_2 + longitude^2\beta_3 + latitude^2\beta_4 + lon * lat\beta_5$$

$$AIC = 1163, BIC = 1201$$

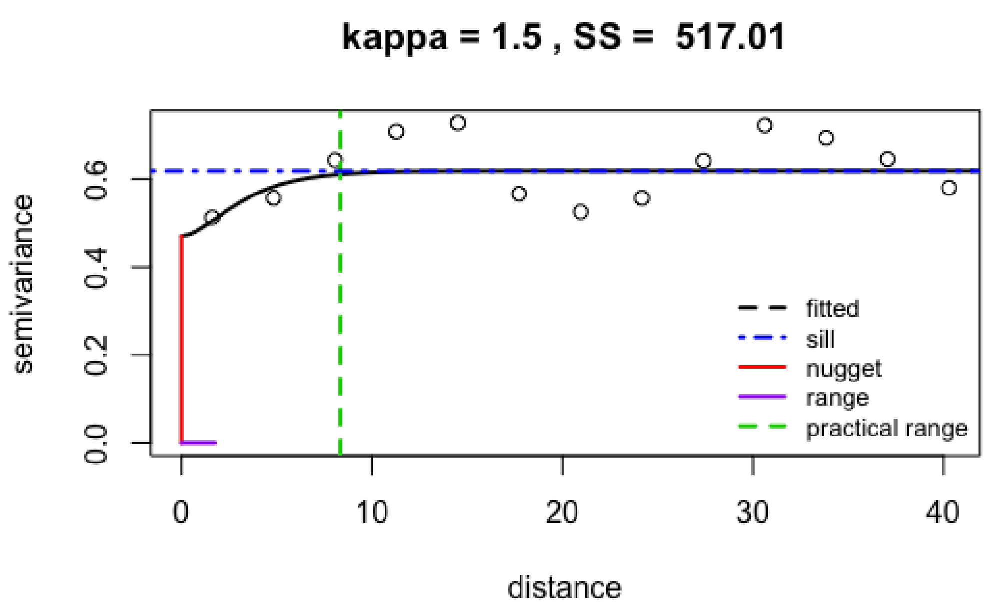


5 Variograms

We obtained the residuals after fitting the trend function and plotted the directional variogram below. There could be an issue because we are using euclidean distance on latitude and longitude over a large area. However the curvature of the earth should not effect regions close to the equator.



From the directional variogram, since none of the lines are vastly different from the rest, we can conclude there is no evidence of possible anisotropies. Next we use least squares to fit the covariogram in the Matèrn family with smoothness equal to (0.5, 1, 1.5, 2.5). The best fit was with a smoothness parameter equal to 1.5 and parameter estimates; $\hat{\tau^2} = 0.4702$, $\hat{\sigma^2} = 0.1476$ and $\hat{\phi} = 1.7581$. We can see the variogram and fitted covariance function below.



6 Model

The model fitted to the data has observation equation

$$X = \mu + \epsilon, \qquad \epsilon \sim N(0, \tau^2 I)$$

and process equation

$$\mu = D\beta + \nu, \qquad \nu \sim N(0, \sigma^2 R(\psi))$$

Where D is the design matrix at each location and $R(\psi)$ is a matrix computed using the Matèrn correlation function. This results in

$$X \sim N(D\beta, \tau^2 I + \sigma^2 R(\psi))$$

For convenience, parametrize as $\gamma^2 = \tau^2/\sigma^2$, so that $X \sim N(D\beta, \tau^2(I + 1/\gamma^2 R(\psi)))$. Then τ^2 can be factorized in the likelihood. Thus our likelihood becomes,

$$L(\beta, \psi, \tau^2, \gamma^2|X) \propto |I + R(\psi)/\gamma^2|^{\frac{-1}{2}} (\tau^2)^{\frac{-n}{2}} \exp\left\{-\frac{1}{2\tau^2}\left[(X - D\beta)'\left(\tau^2\left(I + R(\psi)/\gamma^2\right)\right)^{-1}(X - D\beta)\right]\right\}$$

Sampling of the posterior distribution is achieved by direct sampling of the parameters using the block conditioning,

$$\pi(\beta, \psi, \tau^2, \gamma^2|X) = \pi(\beta|\tau^2, \psi, \gamma^2, X)\pi(\tau^2|\psi, \gamma^2, X)\pi(\psi, \gamma^2|X)$$

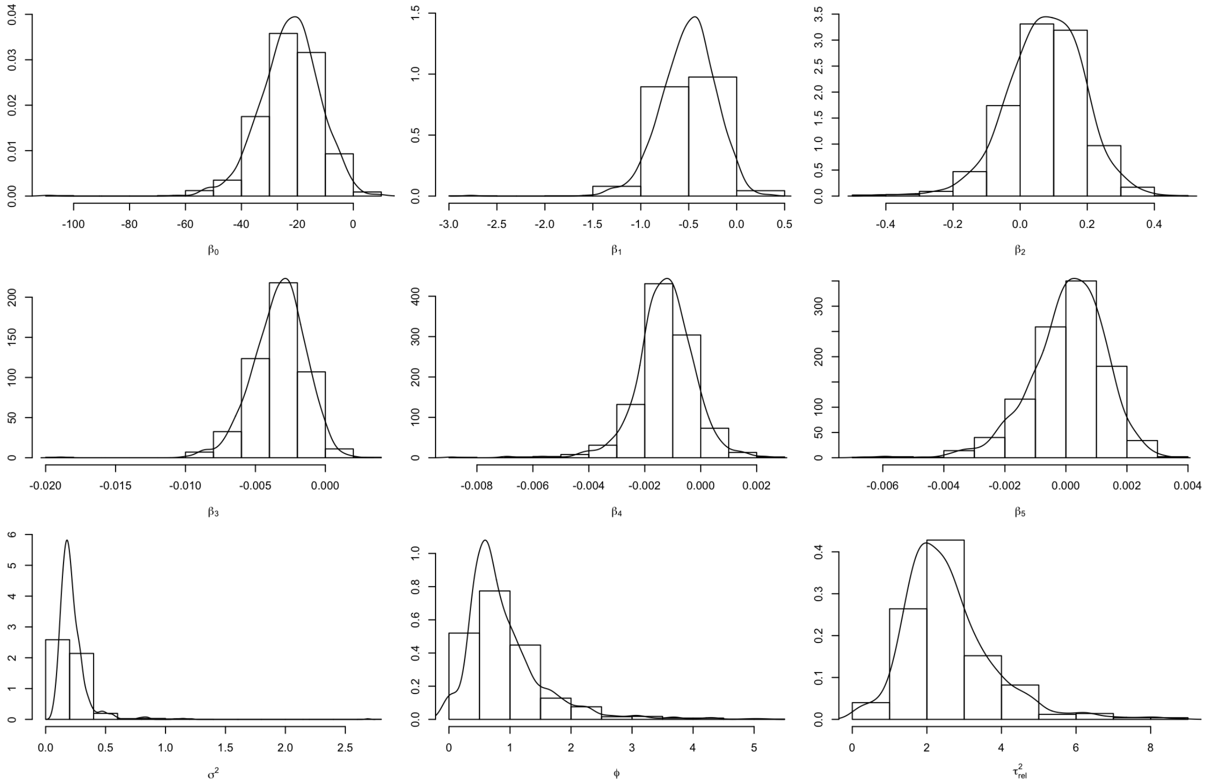
We use a reciprocal prior for σ^2, γ^2, and a squared reciprocal prior for ϕ keeping the optimal smoothness parameter $\kappa = 1.5$.

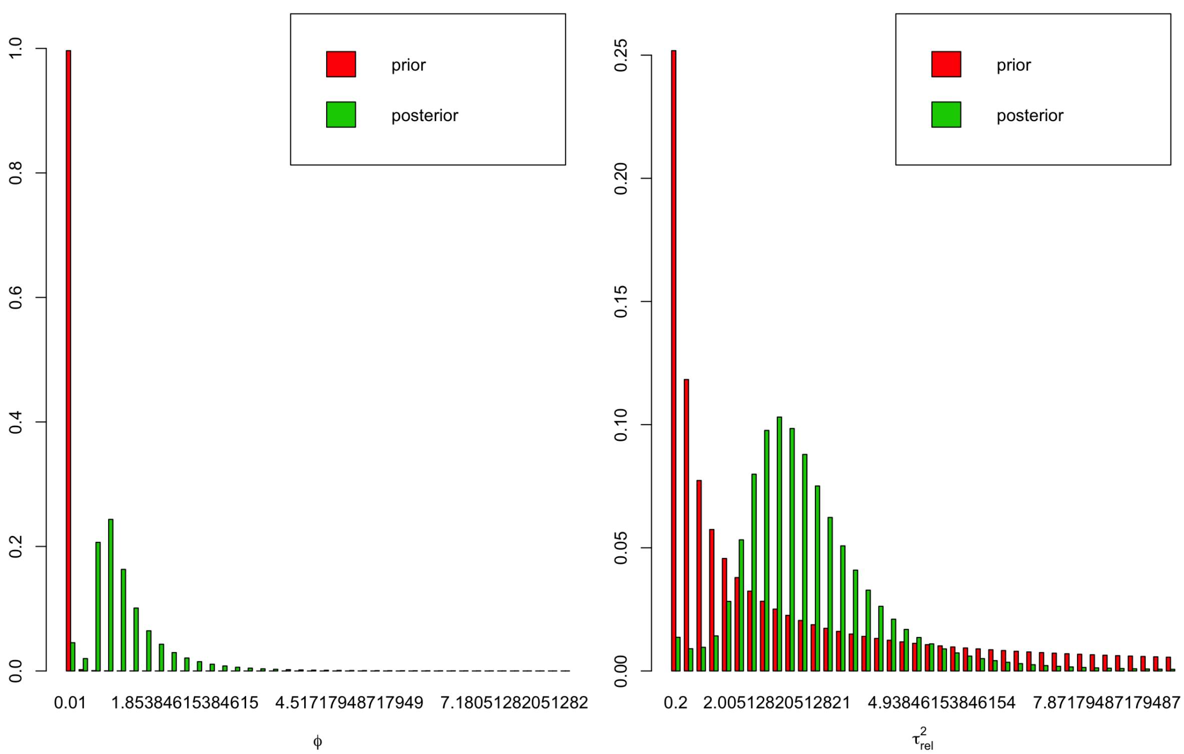
7 Results

The table below contains the parameter estimates for the likelihood fit and the Bayesian model. All of the MLE values are close to the Bayesian estimates. We can see that some of the credible intervals for β contains zero. A better model may only include longitude.

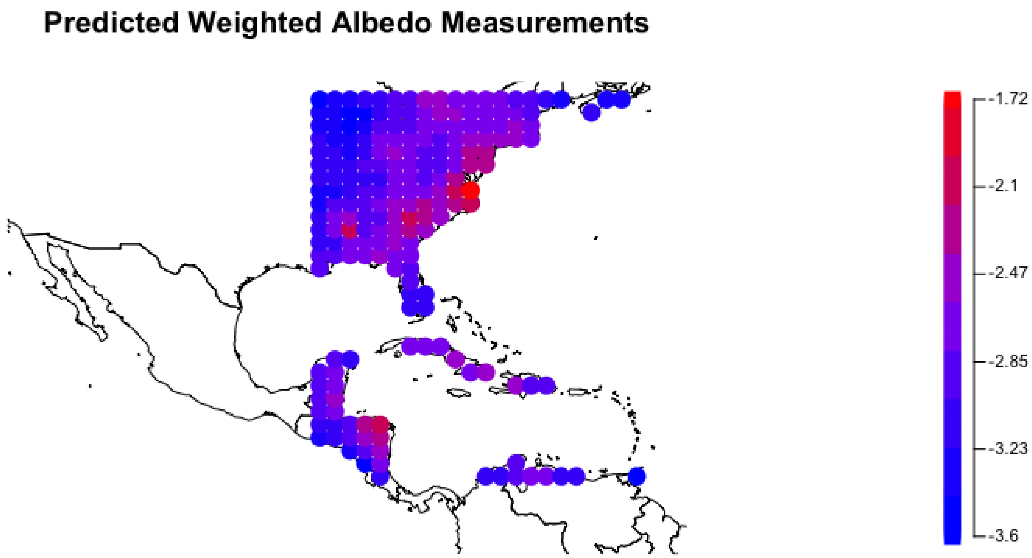
parameter est.	MLE	Bayesian	95% credible interval
$\hat{\beta}_0$ (intercept)	-20.7541	-22.3159	(-44.3305, -3.8998)
$\hat{\beta}_1$ (longitude)	-0.4463	-0.4883	(-1.1018, -0.0071)
$\hat{\beta}_2$ (latitude)	0.1023	0.0778	(-0.1574, 0.2833)
$\hat{\beta}_3$ (longitude^2)	-0.0028	-0.0032	(-0.0072, -0.00003)
$\hat{\beta}_4$ (latitude^2)	-0.0011	-0.0012	(-0.0034, 0.0007)
$\hat{\beta}_5$ (longitude*latitude)	0.0005	0.0002	(-0.0026, 0.0022)
$\hat{\gamma^2} = \frac{\hat{\tau^2}}{\hat{\sigma^2}}$	2.7454	2.4564	(0.6513, 5.6210)
$\hat{\sigma^2}$ (sill)	0.1673	0.1967	(0.0913, 0.5352)
$\hat{\phi}$ (range)	0.5208	0.6246	(0.01, 2.878)

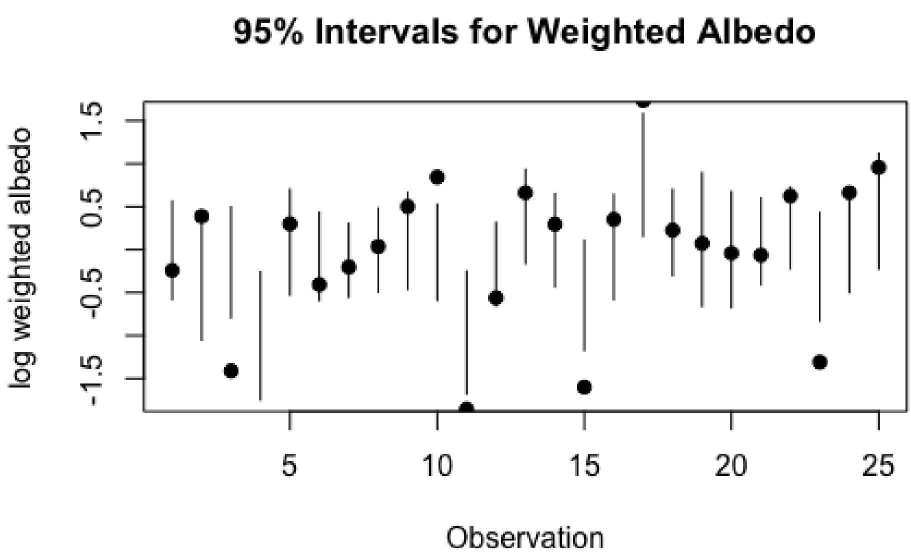
Below are the posterior samples for each parameter.



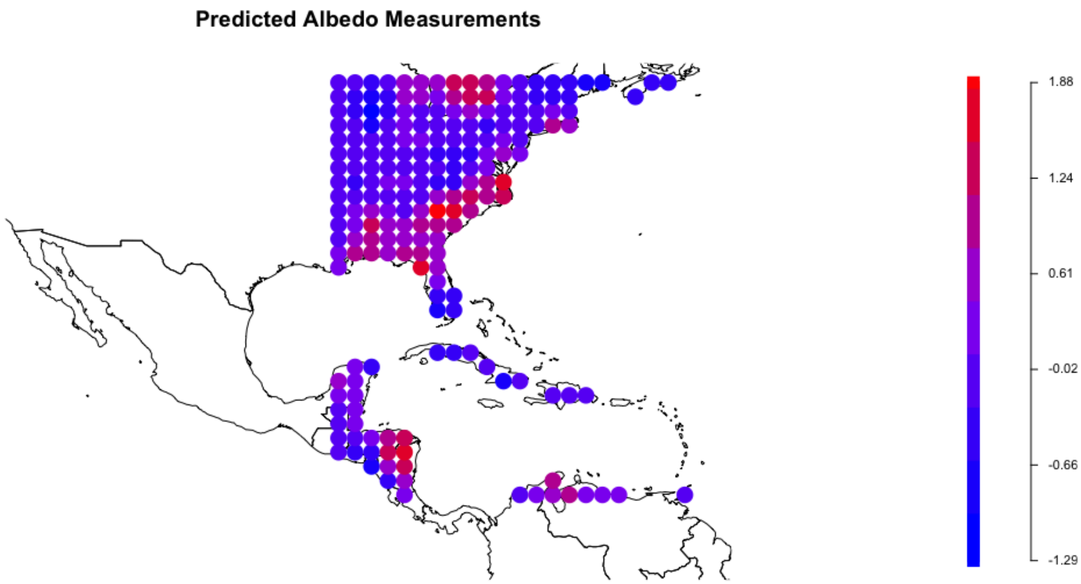


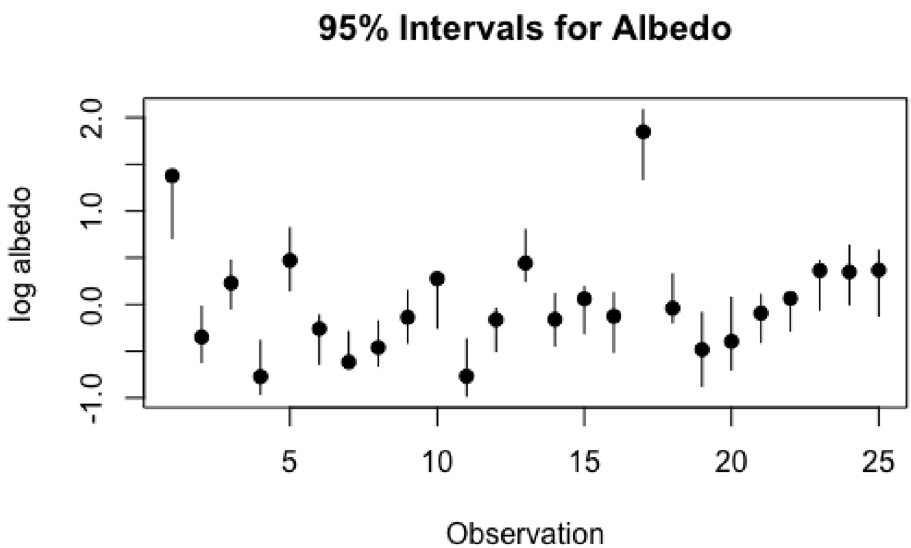
Below we can see the predicted log weighted albedo measurements over the whole area of interest. To check the model, we predict and create 95% credible intervals for each observed data point and find the proportion of data that falls inside of the intervals. Using this method we get 71% accuracy. We can see an example of this will 25 of the 500 intervals below.





Now we predict the non-weighted albedo over the same area and calculate the accuracy. The model was run with the same priors but the optimal smoothness parameter was 2.5. The plots of albedo on the map have similar characteristics. The measurement is higher near the ocean and lower near the center of the US.





All of the observations shown here are contained in the prediction interval but overall, we got an 83.5% accuracy for this model. With more time I would predict over many more points to get a better picture but with the time constraint I had to limit myself to 200 predictions.