Homework 6

Sarah Jarvis
Spatial Statistics AMS 245

June 7, 2022

1 Introduction

We will investigate the model fit of a process convolution with a Markov random field prior. The design matrix will consist of an intercept, latitude, and longitude for each in a blockdiagonal matrix format. This way we can have a set of β coefficients for each satellite. The response variable is a cloglog transformation of albedo because this normalized the data. Figure 1 shows the whole observation data set as well as 1000 randomly selected points for both satellites. The points appear to be representative of the entire data set of albedo. We can see that there is a higher albedo measurement in the central northern area in Texas as well as in the southern region of Mexico near the Yucatan Peninsula. The south eastern region is only shown in GOES 075, so we expect to see discrepancies in the predictions for that domain. Figure 2 illustrates the 1000 transformed albedo measurements. The circles represent the knots used in the process convolution. Figure 2 also shows the distribution of the transformed variable. The cloglog transformation was successful in acheiving normality of the response variable.

Model

Given a set of observations $y_1, ..., y_m$ at locations $s_1, ..., s_m$ a process convolution is defined as

$$y_i = \mu(s) + \sum_{j=1}^{p} k(s - u_j; \omega) w_j + \epsilon_i, \quad \epsilon_i N(0, \tau^2)$$

with priors $w_j \sim N(0, \sigma^2)$, $p(\sigma^2)$, $p(\omega)$ and $p(\tau^2)$. The mean function $\mu(s)$ is the product of a design matrix and some unknown coefficients $X(s)\beta$. The model requires a set of knot points $(u_1, ..., u_m)$, which we choose as 77 equidistant points across the domain of interest. We use a spherical Bezier kernel with the form,

$$k(s; \nu) = \begin{cases} (1 - ||s||^2)^\nu & \text{if } ||s|| < 1, \text{ for } \nu > 0 \\ 0 & \text{otherwise.} \end{cases}$$

To incorporate information from both satellites, we fit the following model

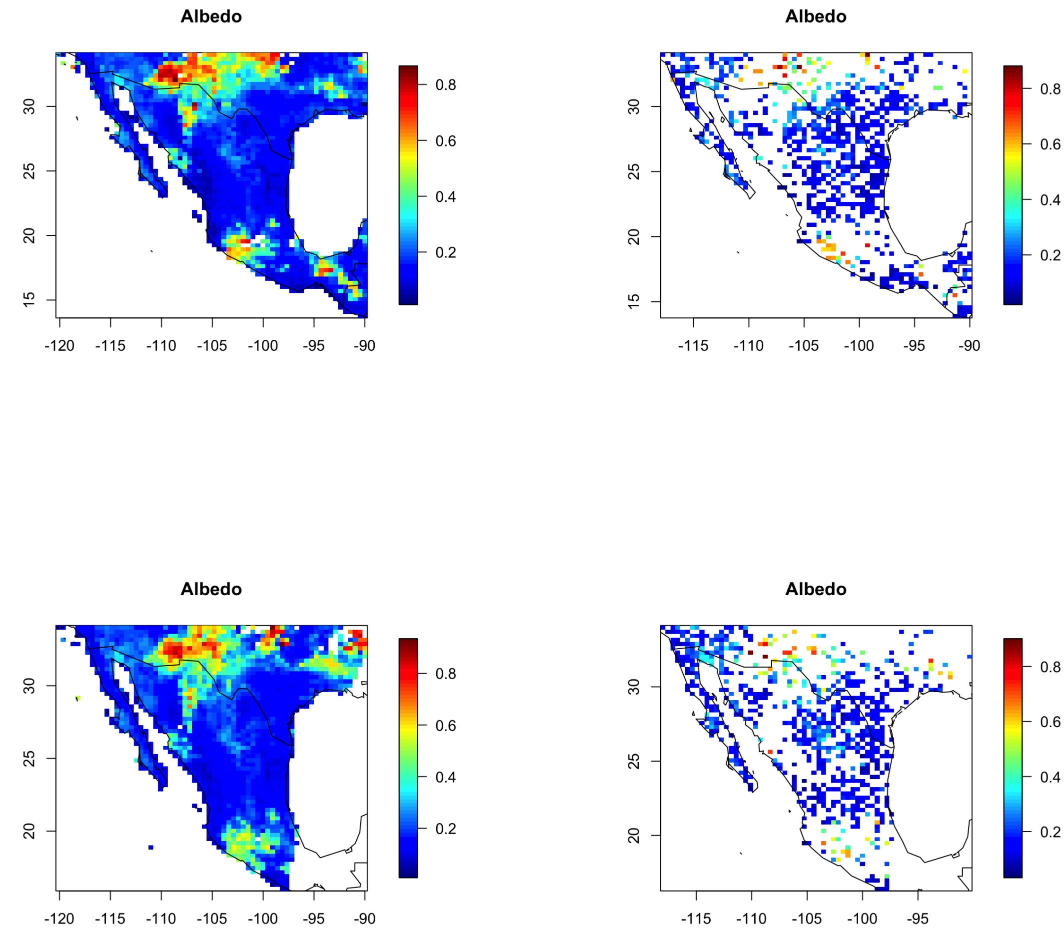


Figure 1: Top: GOES 075 albedo observations and 1000 randomly selected points. Bottom: GOES 135 albedo observations and 1000 randomly selected points.

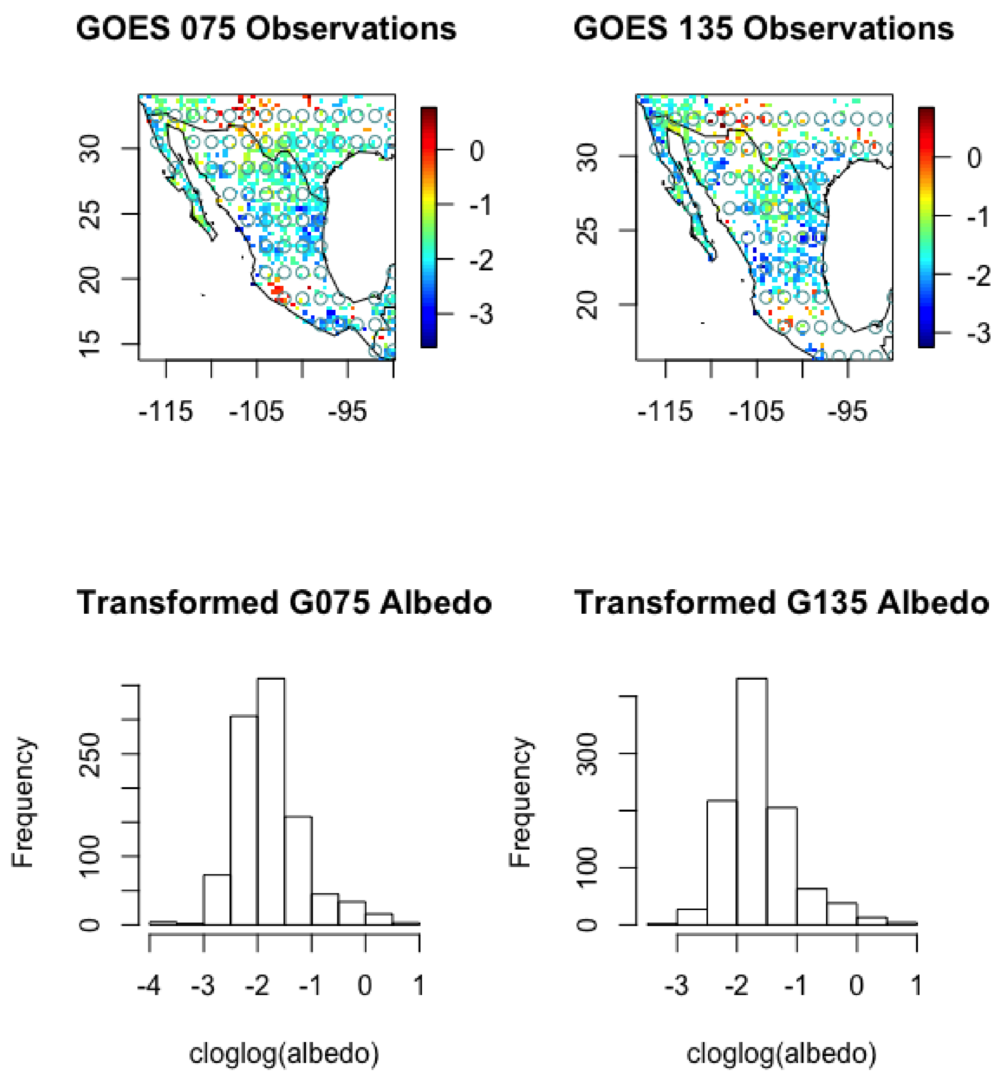


Figure 2: Top: Plot of 1000 randomly selected albedo measurements from each satellite with a cloglog transformation. Circles represent prediction knots. Bottom: Distributions of the transformed albedo response.

$$y_1(s) = x(s)'b_1 + m(s) + e_1(s)$$
$$y_2(s) = x(s)'b_2 + m(s) + d(s) + e_2(s)$$

where $x(s)$ is a set of covariates, m(s) corresponds to the true common albedo, and d(s) to the discrepancy between the two satellites. We write it in vector form as $Y(s) = X(s)B + LM(s) + E(s)$ for an appropriate lower triangular matrix L, and $M(s) = (m(s), d(s))$.

The design matrix will consist of an intercept, latitude, and longitude for each satellite in a blockdiagonal matrix format. This way we can have a set of β coefficients for both GOES 075 and GOES 135. This is important because the satellites are recording the same observations but from different angles and it is possible that one is capturing information that the other is not. If we have two sets of beta coefficients, we can make predictions for each satellite separately. This could be necessary if it was decided that one satellite is more accurate.

Prior Selection

The prior we choose for the β coefficients is $p(\beta) \propto 1$. We choose conjugate inverse gamma priors for τ^2 and ν, $\tau^2 \sim IG(2, 1)$ and $\nu \sim IG(2, 1)$. We place a markov random field (MRF) prior on w, $w \sim GMRF(\lambda W)$. A random vector $x \in \mathbb{R}^n$ is a GMRF with mean μ and precision matrix W iff its density has the form

$$\pi(x) \propto |W|^{1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T W(x-\mu)\right\}$$

$$W_{ij} = \begin{cases} n_i & \text{if } i = j \\ -1 & \text{if } i \sim j \\ 0 & otherwise. \end{cases}$$

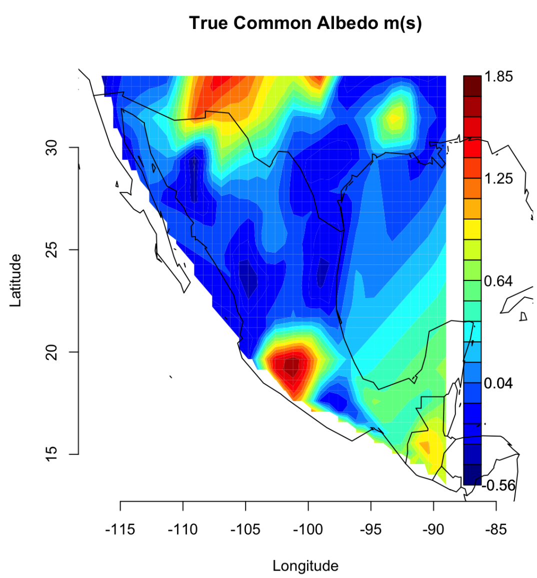
where n_i is the number of neighbors of w_i. The full conditional is a normal distribution and we use a gamma prior for the precision scale $\lambda \sim gamma(0.01, 0.01)$.

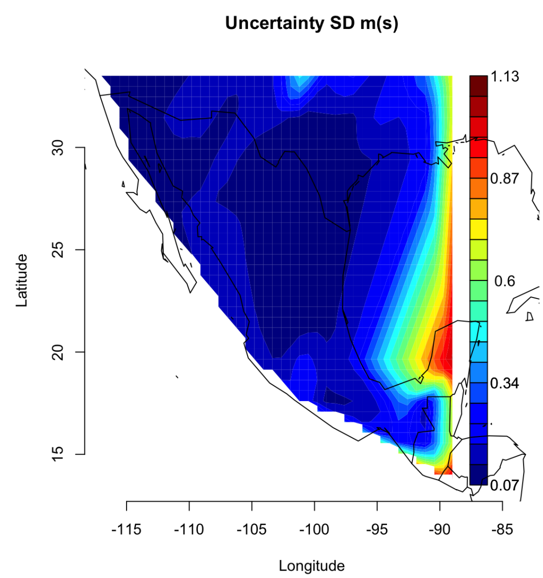
Results

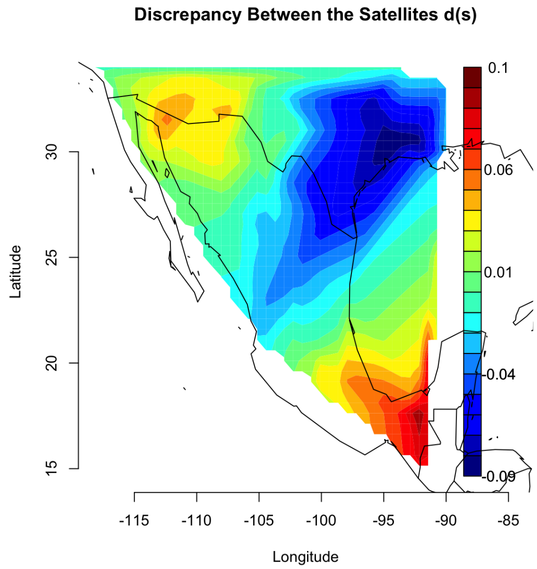
The table below summarizes the posterior distribution of the parameters in the model. The table contains the mean estimate, standard deviation, and lower and upper bounds for a 95% credible interval. All of the coefficients but longitude (for both satellites) do not contain zero and are thus significant. Note that τ^2 is the observational variance, ν is the smoothness parameter in the spherical Bezier kernel, and λ is the precision scale of the convolution coefficients.

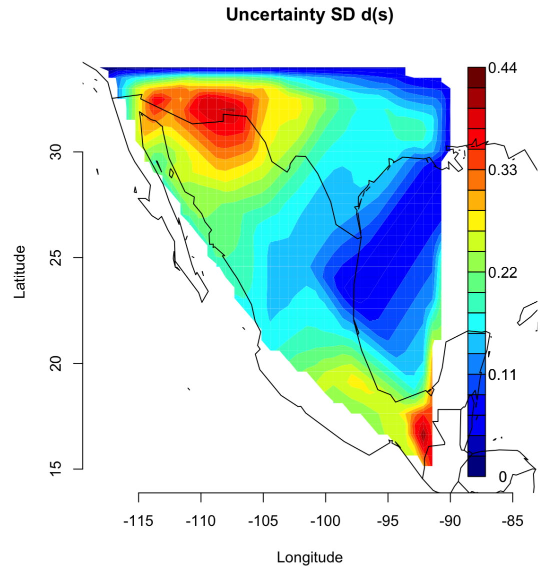
parameter	Mean	SD	CI.lower	CI.upper
intercept_{075}	-2.557	0.399	-3.331	-1.761
lon_{075}	0.002	0.004	-0.006	0.010
lat_{075}	0.043	0.005	0.033	0.054
intercept_{135}	-1.968	0.430	-2.843	-1.145
lon_{135}	0.005	0.005	-0.004	0.014
lat_{135}	0.033	0.007	0.020	0.046
λ	2.613	0.429	1.879	3.549
τ^2	0.204	0.007	0.191	0.217
ν	0.137	0.029	0.087	0.197

Below we can see the predictive surface for the true common albedo m(s) and the discrepancy between the two satellites d(s). We see similar local variability in the predictions to those in the observations. We quantify the uncertainty with the standard deviation of the distributions at each knot. For m(s) there is little uncertainty for most of the region. The higher uncertainty is at the edge where we did not have prediction knots. There is much more uncertainty in the discrepancy d(s) with high SD's around the local variability of d(s). As we expected, there is significant difference in the south east region. This makes sense because we only have data from GOES 075 in that area. There are also large differences around (-95, 30). This we can see when observing differences in the plots of the data above.

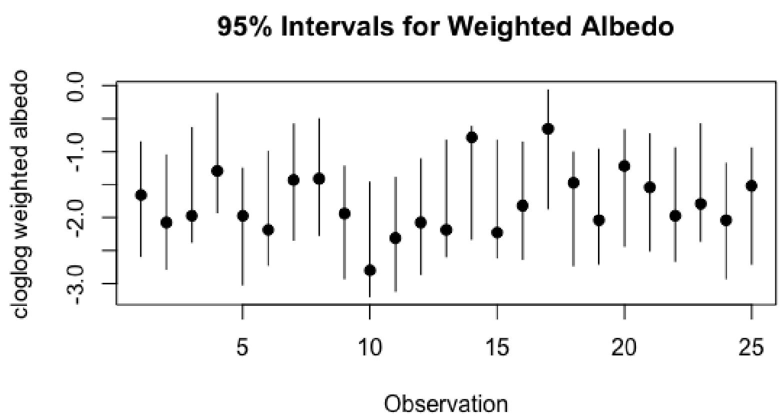


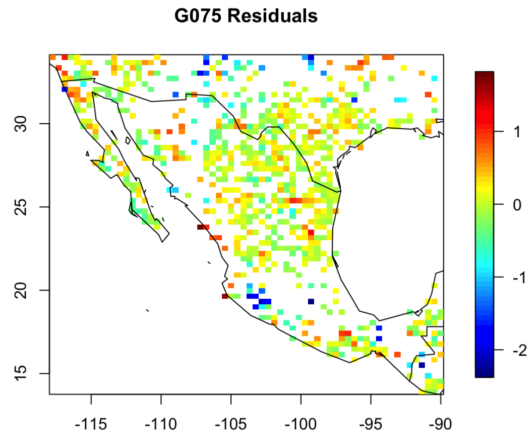


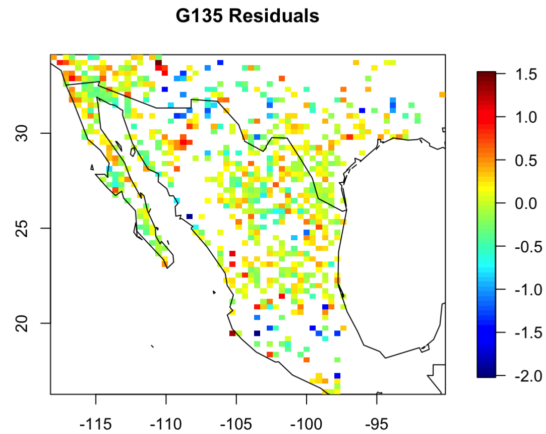




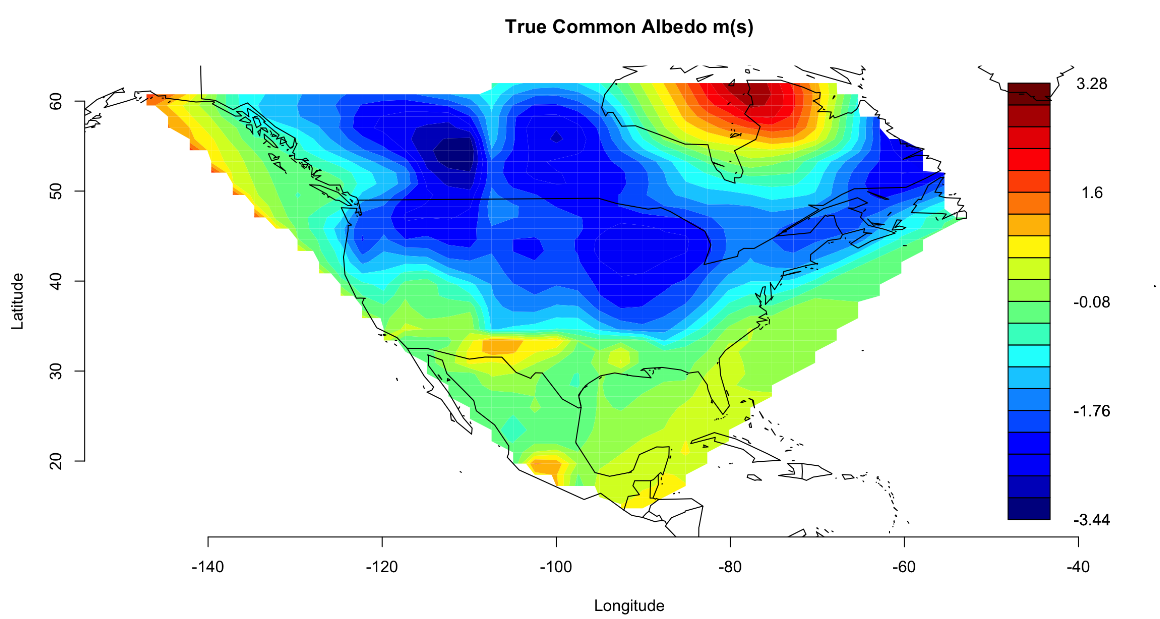
As a measure of accuracy, we make 95% credible intervals for each prediction and calculate the proportion of observations that fall inside of the intervals. Using this approach, we get a 95.4% accuracy. Below we can see 25 of the credible intervals with the observations shown all falling inside. Below the credible intervals is the a spatial representation of the residuals. We can see that a majority of the residuals are close to zero. This shows that our model accurately predicts the data.

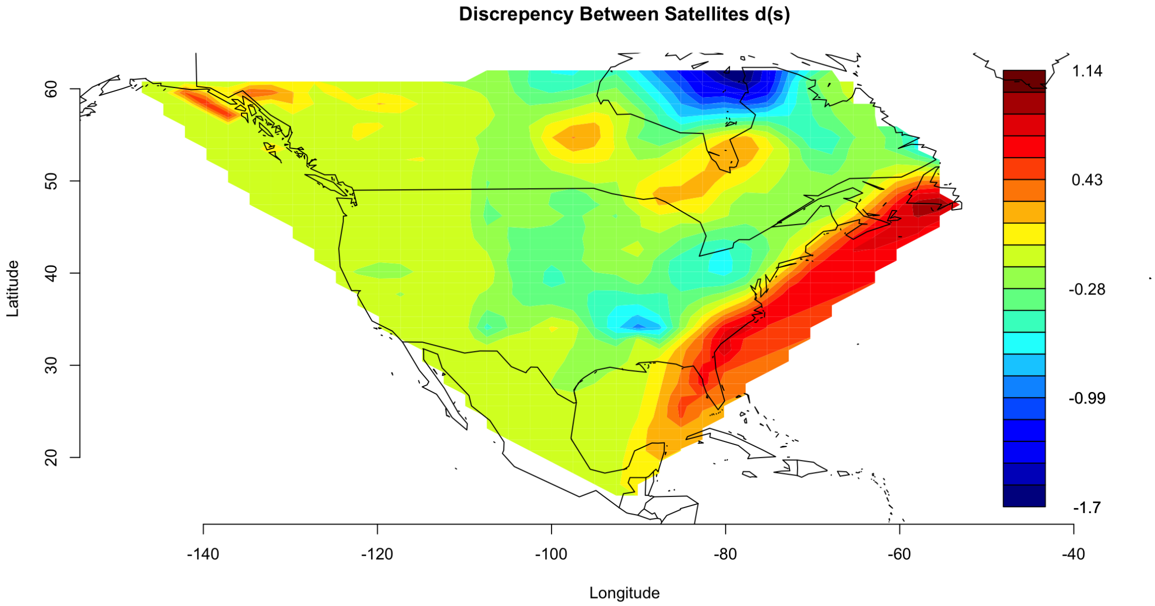






Below is the plot of every region from our class. Please note that it does not make very much sense to combine these results because each model was fit with a different transformation as well as different covariates. The two additional predictions were modeled using a logit transformation. Combining the predictions makes the local variation in Mexico look much less severe.




Discrepency Between Satellites d(s)
Latitude
Longitude
60
50
40
30
20
-140
-120
-100
-80
-60
-40
1.14
0.43
-0.28
-0.99
-1.7