A Spatial Analysis of Colorado School Performance

—

Spatial Statistics Final Project Report

Galen Vincent, Joel Walker

December 10, 2018

Abstract

In this report we analyze a data set containing the performance ratings of each public school in Colorado for the 2016-17 school year. We build a spatial model with predictor variables of the percentage of students on free and reduced lunch and the student teacher ratio at each school, and use kriging with this model to predict the performance of new schools at locations in Colorado. We found that while predicting these performances is possible given that data on free and reduced lunch and student teacher ratio is available, the uncertainty in these predictions is large.

Contents

1 Introduction

The data we analyze in this report was obtained from the US Department of Homeland Security (DHS) and the Colorado Department of Education (CDE). From DHS's public data website, we collected the spatial locations (longitude and latitude) of 1493 public schools in Colorado [1]. From CDE's website, we collected for each public school: their overall performance rating given by CDE [2], the number of students enrolled [3], the percentages of different ethnicities of students enrolled [3], the number of faculty employed [3], the percentage of students eligible for free and reduced lunch (FRL) [3], and the student-teacher ratio (STR) [4]. All of this data came from the 2016-2017 school year. To guide our analysis of this data, we considered three research questions:

1. Can we predict the performance rating of a new school based on their location and the performance of neighboring schools?

2. What factors impact a school's performance rating?

3. In what areas of Colorado should CDE focus their efforts on improving public schools?

As is evident from the above questions, our analysis is focused around each school's performance rating, which is given by CDE on a scale from 0-100. This rating is based on many factors including student standardized test scores, student growth over time, graduation statistics, and many more [2]. Figure 1 shows a spatial map for these performance ratings at each of the 1493 public schools in Colorado for which there is data available.

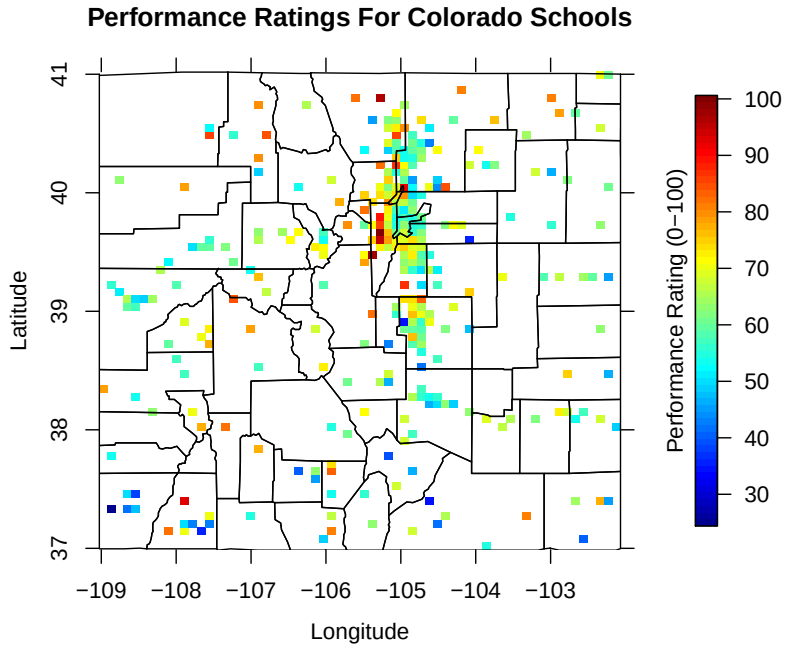
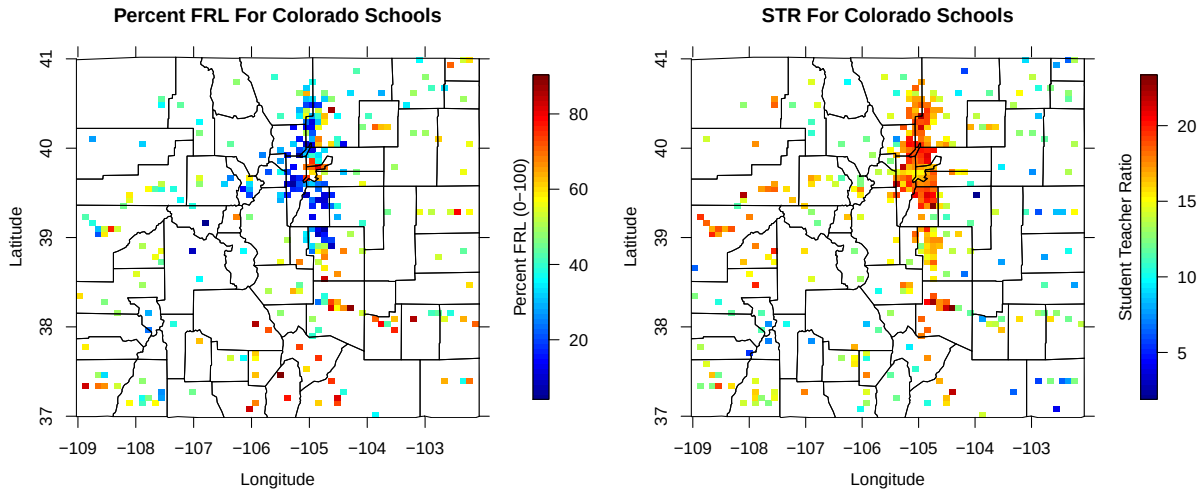


Figure 1: Spatial map of CDE performance ratings at 1493 schools across Colorado. County lines are also shown.

We found from early analysis that the most significant independent variables, or predictors, which influence this performance are FRL[1] and STR. Other independent variables we considered were school size, number of faculty, and percent of different ethnicities in the school's population. These other independent variables were not found to have a significant or independent impact on school performance. Figure 2 shows spatial plots of the significant predictors of FRL and STR across Colorado.



(a) Percentage of students on FRL

(b) Student teacher ratio

Figure 2: Predictors for linear model of school performance. Both (a) percentage of students on free and reduced lunch (FRL) and (b) the schools student teacher ratio (STR).

2 Statistical Methodology

The initial method we used to analyze this data set was simple linear regression of performance as a function of various predictors. The parameters for these models were fit using ordinary least squares minimization. The purpose of this analysis was to determine which of the independent variables in our data set influence a school's performance rating the most. Different predictors were analyzed by looking at p-values, and whole models were compared by looking at their minimized least squares totals.

Once we determined which predictors were most important and had the simple linear model based on these variables, we wanted to ensure that the residuals around this model were weakly stationary (henceforth just called stationary) and normally distributed. We make these assumptions about the residuals as we move along in our analysis, so it is important to check that they are valid assumptions to make. To make sure the residuals are stationary, we needed to confirm that their mean is spatially constant, and that their variogram is solely a function of lag between points. We achieved this by first splitting the data into quadrants based on location and calculating each quadrant's mean and standard deviation. The goal of this was to ensure that neither the mean or covariance of the data

[1]For more information on what qualifies students to be eligible for free and reduced lunch, see `https://www.cde.state.co.us/nutrition/nutrfreeandreducedmaterials`.

is spatially dependent. Next, we estimated a directional variogram to check that covariance between points is only a function of lag, and not direction. To check for normality of the residuals, we created a histogram and normal QQ plot, and inspected them visually.

Once we were confident that the residuals of our simple linear model are stationary and normal, we were able to treat them like a spatial process with the methods we learned this semester. So, the next step was to fit a covariance model to the residuals. We estimated parameters τ, σ, λ, and ν for different covariance models using weighted least squares with the empirical variogram, which was calculated using the robust method proposed by Cressie and Hawkings [5]. These parameters correspond to the nugget, process standard deviation, range, and smoothness parameters, respectively. The covariance models that we fit are shown in Table 1. Once the parameters were fit, we compared the different models by looking at their associated weighted least squares sums.

Table 1: Covariance Model Formulas

Exponential	$C(h) = \sigma^2 \exp(-\frac{h}{\lambda})$
Matérn	$C(h) = \sigma^2 \frac{1-\nu}{\Gamma(\nu)} (\frac{h}{\lambda})^\nu K_\nu(\frac{h}{\lambda})$
Spherical	$C(h) = \begin{cases} \sigma^2 \left[1 - 1.5(\frac{h}{\lambda}) + 0.5(\frac{h}{\lambda})^3\right] & \text{if } h < \lambda \\ 0 & \text{otherwise} \end{cases}$
Cauchy	$C(h) = \sigma^2 \left[1 + (\frac{h}{\lambda})^2\right]^{-\nu}$

While this covariance model estimation for our simple linear model's residuals gave us a good initial feel for the spatial relationships in our data, we eventually moved on to fit the entire model using maximum-likelihood estimation (MLE). We minimized the profiled log likelihood for a spatial process [5] with predictors of FRL and STR to estimate covariance parameters for the Matérn and exponential covariance models. We then used these parameters to find the $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ estimates for our linear model in FRL and STR. These complete spatial models with mean function, spatial covariance process, and white noise process, are better overall models than the simple linear model with empirically estimated covariance function that we first explored. This is because by definition, MLE parameters are the most likely estimates of parameters based on the observed data. We chose between the the Matérn and exponential MLE models using Akaike's information criterion (AIC) and Schwarz's Bayesian information criterion (BIC) [5].

Once we selected a complete spatial model, we then performed kriging in order to predict potential new school's performance ratings. Because there doesn't exist data for FRL and STR of schools which do not exist, it wasn't possible for us to predict a full surface with universal kriging. Instead, we predicted a surface of the residuals around the mean function using simple kriging. Universal kriging would be useful if the FRL and STR of a new school was known, which would be reasonable data to acquire in the first few weeks of a new

school opening. To show proof of concept that our model could be used to predict a school's performance rating in this case, we also performed universal kriging on a few made up data points.

3 Results

The school performance data by itself is not inherently stationary or normally distributed. This is evident from Figure 1. Schools at lower latitudes seem to have much lower average scores than school at higher latitude. There are also an abundance of other factors that play in to a school's performance rating besides its spatial location. It turns out that this data isn't normally distributed either. We don't show the histogram or normal QQ plot here, but the performance data obviously did not follow a normal distribution. In order to make our data stationary and normal to enable the use of our statistical methodology, we decided to fit a linear model for school performance of

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \tilde{\varepsilon}, \tag{1}$$

where z is a given schools performance rating, $\tilde{\varepsilon}$ is Gaussian distributed white noise with mean zero and variance $\tilde{\sigma}^2$, and the covariates $\mathbf{x} = (x_1, x_2)$ are given by

$$\mathbf{x} = (\text{FRL},\ \text{STR}),$$

where FRL is the percent of students at the given school on a free and reduced lunch plan, and STR is the student teacher ratio at the given school. Spatial plots of these covariates are shown in Figure 2. Other predictor variables were explored, but these are the two that ended up having the most significant and independent impact on school performance. We first predicted the parameters for model (1) using ordinary least squares (OLS). These estimates, standard errors, and parameter p-values are shown in Table 2. Note that these are OLS estimates, so the $\hat{\tilde{\sigma}}^2$ estimate may be slightly different than the MLE estimate of the same parameter, but the OLS estimate is unbiased. Note that the p-values for the $\boldsymbol{\beta}$ parameters are all smaller than 0.05, indicating that they are indeed significant predictors.

Table 2: OLS Parameter Estimates for Model (1)

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\tilde{\sigma}}^2$
Estimate	80.525	-0.320	-0.176	12.27
Std. Err.	0.991	0.012	0.039	-
p-value	2×10^{-16}	2×10^{-16}	7.32×10^{-6}	-

Figure 3 shows a spatial plot of the residuals from this linear model. The goal of this model was to make the data stationary and normally distributed. To test for stationarity, we split the residuals up into quadrants as shown in Figure 3, and looked at the mean and standard deviation of the data in each quadrant. The results are shown in Table 3.

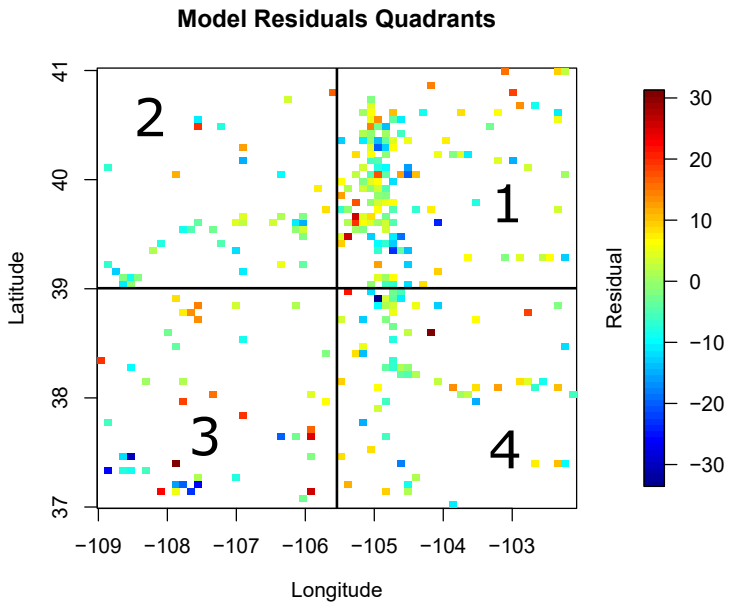


Figure 3: Model residuals with the quadrants used to split the data to check for stationarity.

Table 3: Mean and Standard Deviation of Data in Quadrants from Figure 3

Quadrant	1	2	3	4
Mean	1.04	-2.49	-0.45	0.68
Std. Dev.	11.94	11.90	14.84	12.36

Our conclusion from Table 3 is that the means of the four quadrants are close compared to the $(-30, 30)$ range of the data itself, and the story is similar with the standard deviations. This indicates that the residual process has constant mean and variance across its domain. To check that the residual process is isotropic, i.e. to check that it's covariance function only depends on lag, h, we performed a directional variogram at four different angles. The results are shown in Figure 4. From this figure, it is evident that the covariance structure of the process stays relatively consistent regardless of angle, which means that we can assume the residual data is indeed isotropic. Based on the fact that these residuals have constant spatial mean and variance, and that the covariance model is a function of only lag h, we conclude that our data is weakly stationary.

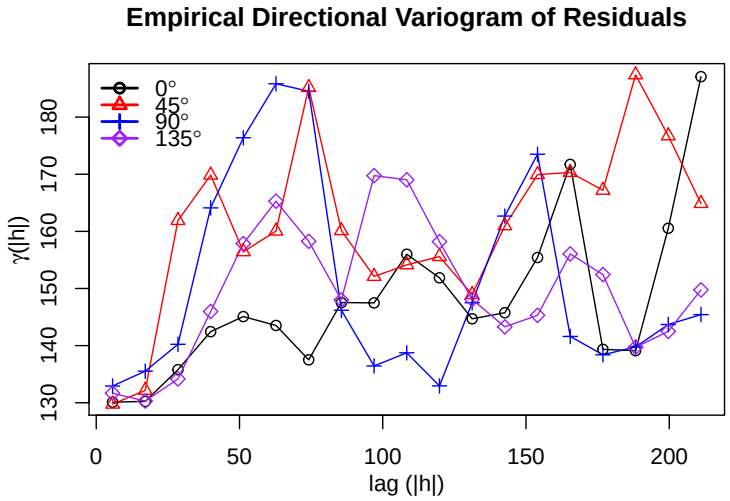
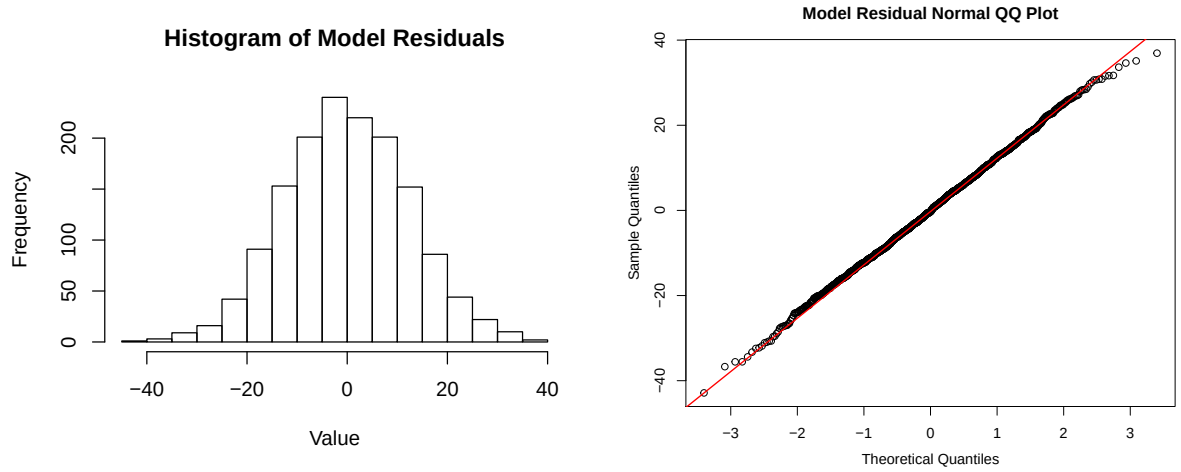


Figure 4: Directional variogram of model residuals in four different directions. Lag is in miles. No major changes by direction indicates that residuals are isotropic.

Next, we showed that the residuals of the linear model are normally distributed. Figure 5 shows a histogram and normal QQ plot of the residuals. It is evident visually from these figures that our residuals are indeed normally distributed. Because we have showed that the residuals are both weakly stationary and normally distributed, we can continue on to apply the spatial statistics methodology we have learned in this class.



(a) Histogram

(b) Normal QQ plot

Figure 5: Both (a) histogram and (b) normal QQ plot of the model residuals. Both indicate that the residuals are normal.

At this point in the analysis, we have shown that we can work with the residuals as a stationary, isotropic, normally distributed process of their own. So next, we attempt to find a spatial model for the residuals of model (1) as a process of their own. From model (1), we can model the residuals $\tilde{\varepsilon}$ as

$$\tilde{\varepsilon}(\boldsymbol{s}) = Y(\boldsymbol{s}) + \varepsilon, \tag{2}$$

where $Y(\boldsymbol{s})$ is a stationary zero mean random process with covariance function $C(h)$, and ε is a white noise process with mean zero and variance τ^2. Because $\tilde{\varepsilon}$ are residuals themselves, we know that the whole process has mean zero, so there is no μ component to this process. Adding this into the simple linear model (1), our entire spatial model now becomes

$$z(\boldsymbol{s}) = \beta_0 + \beta_1 x_1(\boldsymbol{s}) + \beta_2 x_2(\boldsymbol{s}) + Y(\boldsymbol{s}) + \varepsilon. \tag{3}$$

Next, we estimated the covariance function $C(h)$ for $Y(\boldsymbol{s})$ by fitting different variogram models to empirical variogram estimates using Cressie weighted least squares [5]. Figure 6 shows the robust variogram estimate [5] with the different models we attempted to fit. The formulas for the different covariance models associated with these variograms are shown in Table 1. The parameter estimations, including the nugget term τ^2, and value of each weighted least squares sum are shown in Table 4.

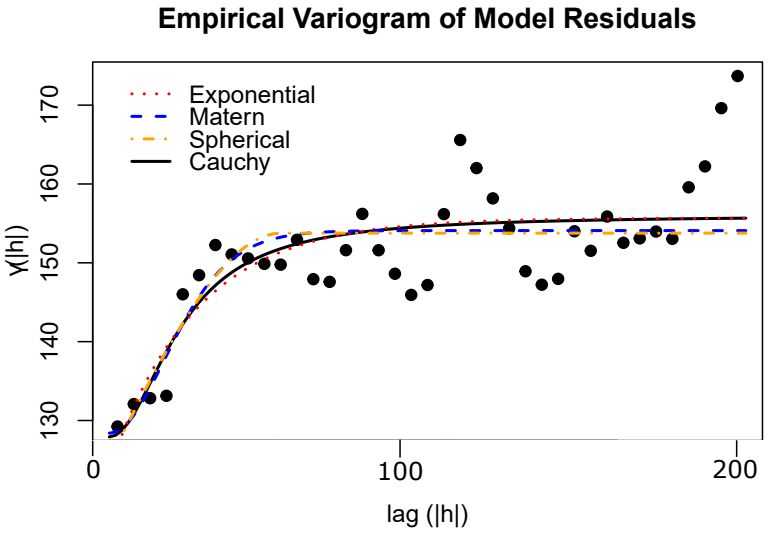


Figure 6: Robust empirical variogram with weighted least squares fits for different covariance models. Lag in miles.

Table 4: Weighted Least Squares Fits for Robust Variogram Estimate

	$\hat{\tau}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\nu}$	WLS Value
Exponential	11.147	5.418	27.420	-	1176.978
Matérn	11.339	4.865	1.212	125.5	1039.817
Spherical	11.166	4.865	53.459	-	1109.995
Cauchy	11.166	4.865	53.459	0.5	1038.637

Based on the WLS values from Table 4, we originally concluded that the Cauchy covariance model was the best fit for our data. That is, until we learned how to do maximum likelihood estimation of our entire process.

Once we learned about MLE in class, we used it to estimate the parameters for all of model (3) as described in Section 2. Running MLE optimization with our data took quite a while (an hour or so for each model), so we stuck to the simple models of Exponential and Matérn. Even though WLS told us that Cauchy would be the best, we wanted to go with more simple, well understood models if we were going to spend so much time doing MLE on them. We then compared these models using AIC and BIC [5]. The MLE parameters and associated AIC and BIC values are compared in Table 5. The associated estimate standard errors are shown in Table 6.

Table 5: MLE Parameter Estimates

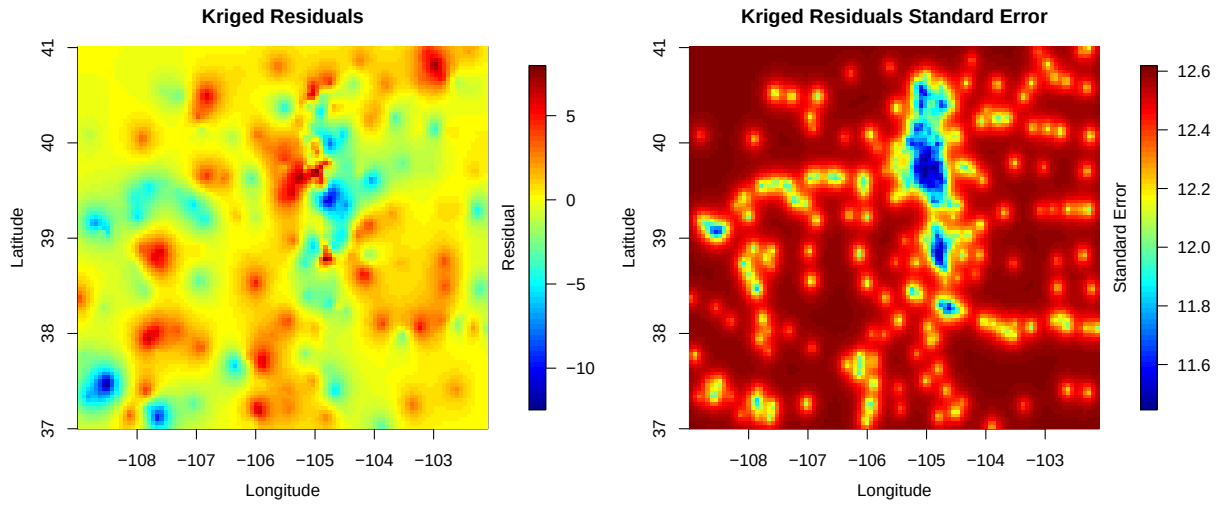
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\tau}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\nu}$	AIC	BIC
Exponential	81.03	-0.343	-0.173	11.20	5.79	9.84	-	11658.64	11695.80
Matérn	81.03	-0.343	-0.173	11.19	5.79	10.17	0.476	11660.64	11703.11

Table 6: MLE Parameter Estimate Standard Errors

Std. Errs.	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\tau}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\nu}$
Exponential	1.244	0.0157	0.0391	0.246	0.668	3.13	-
Matérn	1.245	0.0157	0.0391	0.294	0.690	6.17	0.342

Based on AIC and BIC, we chose the exponential covariance function, which is interesting, as it is actually the worst model according WLS (see Table 4). This speaks to the difference that MLE model fitting can have versus weighted least squares based on empirical variogram.

Based on model (3) with exponential covariance function and parameter estimates shown in Table 5, we performed two different types of kriging. First, we predicted a surface of the residual process $Y(\boldsymbol{s}) + \varepsilon$ around the mean function. This predication surface was calculated with simple kriging on a 100×100 grid of locations across Colorado. Simple kriging was used because we know that the residual process mean function is zero. The predicted surface, along with the associated standard error at each point, is shown in Figure 7. Note that the uncertainty is so large in these plots because the nugget term τ^2 is large for our data.



(a) Predicted residuals from kriging

(b) Standard error of predictions

Figure 7: Simple kriging surface of residuals around the model based on MLE parameters. Both (a) the predicted values and (b) the standard errors for these values.

There are some interesting conclusions to be drawn from Figure 7. First is that there are definite pockets of higher and lower variation around our mean model. A specific area of interest is in the Denver area (around lon = -105, lat = 40), where NW of Denver seems to perform much better than the mean (red in color), while SE of Denver seems to perform much below the mean (blue in color). Even with relatively high uncertainty, the discrepancy between these areas is significant. This indicates that there is spatial dependence on school performance, even once a mean function based on FRL and STR is subtracted out. Predicting behavior around the mean function in not so densely populated areas of Colorado proves to be more difficult, with high uncertainty and the kriging predictions trending to zero due to the nugget term.

The second type of kriging we performed on the data was a proof of concept for how our model might be used in real life. The mean function depends on FRL and STR, which are variables that are not readily available at every location across Colorado. However, if a new school were to open, they could get an idea for these predictors quite early in the school's life, and then predict the future performance of the school based on a universal kriging prediction. Universal kriging takes into account the non-constant mean function of model (3), so we can do universal kriging at some location as long as we have the information about the predictors. To show this functionality, we made up four data points with predictor data and performed universal kriging at these locations. The results are shown in Table 7.

Table 7: Universal Kriging Predictions at Made up Points

Longitude	Latitude	FRL	STR	Universal Kriging Prediction	Std. Err.
-104.4	39.75	20	20	67.57	12.01
-108.5	41.00	70	5	56.13	12.64
-105.0	38.00	10	10	76.77	12.35
-107	39.50	50	50	54.13	12.46

4 Discussion

The first research question was to predict the performance of new schools based on their location and the performance of neighboring schools. We found that the performance of new schools can indeed be predicted based on their location, given that the FRL and STR data is known, using the method based on universal kriging and the spatial model which we detailed in Section 3. However, there is a significant amount of uncertainty associated with the predictions from this model. This is evident both in Figure 7 and in Table 7. There are areas of slightly lower uncertainty near areas with higher school density, like around Denver and Colorado Springs, but it is still higher than we would like it in these places. This high uncertainty is due to the data's large nugget term. The large nugget term indicates that there is a lot of extra randomness in a school's performance rating than can be captured by just FRL, STR, and location. In reality, there are likely many factors that play a part in a schools overall performance rating, and though we took a couple of the most important factors into account, it would be very difficult to build a model that takes care of all of the contributing factors.

The second research question was to explore which factors impact a school's performance rating. We found that the most important factors are percentage of students eligible for free and reduced lunch and student-teacher ratio. The linear model described in Table 2 shows statistically significant influence of these two factors on school performance rating. If we were to spend more time looking into possible predictors, we would have looked at average household income around each of the schools, education levels of students' parents, or other potentially important predictors.

The third research question was to find the areas where the Colorado Department of Education should focus their efforts on improving public schools. Based on this research, it is difficult to provide any specific area that the CDE should focus their improvement efforts. Though we showed that the performance of schools close to one another are spatial correlated, performance itself is not inherently spatially dependant, and there is much uncertainty in predicting performance based on location. The more important predictors of a school's performance rating is their FRL and STR. However, when taking FRL and STR into account, there are still areas where schools are performing below average. This is indicated by the blue spots in Figure 7. CDE could invest into more research in these areas to figure out why they are performing poorly compared to the mean even when FRL and STR are taken into account. There may be other significant factors which impact school performance in these locations.

There are additional questions that this research has brought up for our group. Specifically, we are curious about district level performance data as opposed to that of individual schools. Looking at district level data might reduce the noise that we saw while analyzing the school data, but it would drastically change our statistical model as the data would exhibit discrete spatial variation as opposed to continuous.

References

[1] Homeland Infrastructure Foundation-Level Data (2018). Public Schools. URL: https://hifld-geoplatform.opendata.arcgis.com/datasets/public-schools.

[2] Colorado Department of Education (2017). CDE 2017 School Plan Type Assignments 2010-2017. *Colorado Department of Education Performance Framework Results*. URL: https://www.cde.state.co.us/accountability/performanceframeworkresults.

[3] Colorado Department of Education (2017). 2016-17 Pupil Membership by School, Ethnicity, Gender, and Grade. *Colorado Department of Education Pupil Membership - School Data*. URL: https://www.cde.state.co.us/cdereval/pupilcurrentschool.

[4] Colorado Department of Education (2017). Student Teacher Ratios. *Colorado Department of Education School/District Staff Statistics*. URL: https://www.cde.state.co.us/cdereval/staffcurrent.

[5] Bandyopadhyay, Soutir (2018). MATH 432/532: Spatial Statistics Notes.

Appendix

```
schools <- fread('CO_Schools_Data_Final.csv',drop=1)

# performance from latitude
model <- lm(schools$PERFORMANCE~schools$LATITUDE)
plot(schools$LATITUDE, schools$PERFORMANCE,xlab="Latitude",ylab="Performance")
plot(model)
summary(model)

# performance from latitude and longitude
model <- lm(schools$PERFORMANCE~schools$LATITUDE+schools$LONGITUDE)
plot(model)
summary(model)

# performance from STR
model <- lm(schools$PERFORMANCE~schools$STR)
plot(schools$PERFORMANCE~schools$STR,xlab="Student Teacher Ratio",ylab="Performance")
cut_outlier <- schools[schools$STR < 40,]
plot(cut_outlier$PERFORMANCE~cut_outlier$STR,xlab="Student Teacher Ratio",ylab="Performance")


# performance from FRL
model <- lm(schools$PERFORMANCE~schools$FRL)
plot(schools$PERFORMANCE~schools$FRL,xlab="% Free and Reduced Lunch Eligible",ylab="Performance")
abline(model, col="red", lwd=2)
plot(model)
summary(model)

# perfomance from percent white students
model <- lm(schools$PERFORMANCE~schools$WHITE)
plot(schools$PERFORMANCE~schools$WHITE,xlab="% White",ylab="Performance")
abline(model, col="yellow", lwd=2)
plot(model)
summary(model)

# performance from enrollment
model <- lm(schools$PERFORMANCE~schools$ENROLLMENT)
plot(schools$PERFORMANCE~schools$ENROLLMENT)
abline(model, col="blue", lwd=2)
```

```
plot(model)
summary(model)

# performance from STR and FRL
library(fields)
model <- lm(schools$PERFORMANCE~schools$FRL+schools$STR)
plot(model$residuals,ylab="residual")
plot(model)
summary(model)

schoolsn <- schools[!is.na(schools$PERFORMANCE)]
schoolsn <- schoolsn[!is.na(schoolsn$STR)]
schoolsn <- schoolsn[!is.na(schoolsn$FRL)]

library(fields)
library(ggplot2)
library(ggmap)
library(maps)
library(mapdata)
library(data.table)

states <- map_data("state")
co_df <- subset(states, region == "colorado")
counties <- map_data("county")
co_county <- subset(counties, region == "colorado")

counties <- map_data("county")
co_county <- subset(counties, region == "colorado")
co_county$group <- factor(co_county$group)
class(co_county$group)
co_new <- split(co_county,co_county$group)

par(mar = c(4.1,4.1,4.1,4.1),mgp=c(2.5,1,0))
quilt.plot(schoolsn$LONGITUDE,schoolsn$LATITUDE,model$residuals,main = "Model Residuals",xlab = "Longitude",ylab = "Latitude",ax
axis(1, seq(-109,-103,1), col = NA, col.ticks = 1)
axis(3,col= NA,labels=F,col.ticks=1)
axis(2, col = "black", col.ticks = 1)
axis(4,col="black",labels=F)
for(i in (1:length(co_new))){
  co_new[[i]] <- rbind(co_new[[i]],co_new[[i]][1,])
  polygon(co_new[[i]]$long,co_new[[i]]$lat)
}
mtext("Residual", 4, line=1.5)

qqnorm(model$residuals,main="Model Resiudal Normal Q-Q Plot")
qqline(model$residuals,col="red",lwd=1.5)

# Determine stationarity of residuals
library(fields)
library(data.table)

schools <- fread('CO_Schools_Data_Final.csv',drop=1)

schoolsn <- schools[!is.na(schools$PERFORMANCE)]
schoolsn <- schoolsn[!is.na(schoolsn$STR)]
schoolsn <- schoolsn[!is.na(schoolsn$FRL)]

B0 <- 81.0335820
B1 <- -0.3432731
B2 <- -0.1730419

res <- schoolsn$PERFORMANCE - (B0 + B1*schoolsn$FRL + B2*schoolsn$STR)
xy <- cbind(schoolsn$LONGITUDE, schoolsn$LATITUDE)

par(mfrow = c(1,1),mar = c(4.25,4.25,4,4), mgp = c(2.75,1,0))
quilt.plot(xy[,1],xy[,2],res, main = "Model Residuals Quadrants", xlab = "Longitude", ylab = "Latitude")
mtext("Residual", 4, line=1)
abline(h = ysplit,lwd = 2)
```

```
abline(v = xsplit,lwd = 2)

# Break up residuals into sub grid based on location
ngrid <- 4

xmin <- min(xy[,1])
xmax <- max(xy[,1])
ymin <- min(xy[,2])
ymax <- max(xy[,2])

xsplit <- (xmin + xmax)/2
ysplit <- (ymin + ymax)/2

div <- sqrt(ngrid)

div.list <- list()
for(i in 1:4){
  div.list[[i]] <- list()
}
count <- c(1,1,1,1)

for(i in 1:length(res)){
  if(xy[i,1] < xsplit && xy[i,2] < ysplit){
    div.list[[3]][[count[3]]] <- res[i]
    count[3] <- count[3] + 1
  }else if(xy[i,1] > xsplit && xy[i,2] < ysplit){
    div.list[[4]][[count[4]]] <- res[i]
    count[4] <- count[4] + 1
  }else if(xy[i,1] > xsplit && xy[i,2] > ysplit){
    div.list[[1]][[count[1]]] <- res[i]
    count[1] <- count[1] + 1
  }else if(xy[i,1] < xsplit && xy[i,2] > ysplit){
    div.list[[2]][[count[2]]] <- res[i]
    count[2] <- count[2] + 1
  }
}

div <- list()
for(i in 1:4){
  div[[i]] <- unlist(div.list[[i]])
}

means <- sapply(div,mean)
sds <- sapply(div,sd)

# Variogram for the data
library(data.table)
library(fields)
library(geoR)

schools <- fread('CO_Schools_Data_Final.csv',drop=1)

model <- lm(schools$PERFORMANCE~schools$FRL+schools$STR)
plot(model$residuals,ylab="residual")
#plot(model)
summary(model)
hist(model$residuals,main = "Histogram of Model Residuals",xlab = "Value")

a<-shapiro.test(model$residuals)

schoolsn <- schools[!is.na(schools$PERFORMANCE)]
schoolsn <- schoolsn[!is.na(schoolsn$STR)]
schoolsn <- schoolsn[!is.na(schoolsn$FRL)]
res <- model$residuals

quilt.plot(schoolsn$LONGITUDE,schoolsn$LATITUDE,res)

#export residual data
```

```
#toexp <- as.data.frame(cbind(schoolsn$LONGITUDE,schoolsn$LATITUDE,res))
#colnames(toexp) <- c("LONG","LAT","RESID")
#fwrite(toexp, "residuals.csv")

dist.mat <- rdist(data.frame(lon = schoolsn$LONGITUDE,lat = schoolsn$LATITUDE))
dist.max <- max(dist.mat)/2
vario <- variog(coords = cbind(schoolsn$LONGITUDE,schoolsn$LATITUDE),data = res, option = "bin",breaks = seq(0,dist.max,l=40))
vario$n
par(mgp = c(2,1,0))
plot(vario$u,vario$v,pch = 19,xlab = "lag (|h|)",ylab = expression(paste(gamma,"(|h|)")),main = "Emprical Variogram of Model Res

vario.d <- variog4(coords = cbind(schoolsn$LONGITUDE,schoolsn$LATITUDE),data = res, option = "bin",breaks = seq(0,3,l=20))
vario.d$`0`$n
vario.d$`45`$n
vario.d$`90`$n
vario.d$`135`$n

vario.d2 <- variog4(coords = cbind(schoolsn$LONGITUDE,schoolsn$LATITUDE),data = res, option = "bin",breaks = seq(0,3,l=20))

par(mar = c(3.5,3.5,3.5,3),mgp = c(2,1,0))
plot(vario.d$`0`$u,vario.d$`0`$v,type="l",xlab = "lag (|h|)",ylab = expression(paste(gamma,"(|h|)")), main = "Emperical Directio
lines(vario.d$`45`$u,vario.d$`45`$v,col="red")
lines(vario.d$`90`$u,vario.d$`90`$v,col="blue")
lines(vario.d$`135`$u,vario.d$`135`$v,col="purple")
points(vario.d$`0`$u,vario.d$`0`$v,pch=1,col="black")
points(vario.d$`45`$u,vario.d$`45`$v,pch=2,col="red")
points(vario.d$`90`$u,vario.d$`90`$v,pch=3,col="blue")
points(vario.d$`135`$u,vario.d$`135`$v,pch=5,col="purple")
legend(0.1,188,legend = c(expression(paste("0",degree)),
                          expression(paste("45",degree)),
                          expression(paste("90",degree)),
                          expression(paste("135",degree))),
       col=c("black","red","blue","purple"),pch = c(1,2,3,5),lty=c(1,1,1,1),lwd=c(2,2,2,2),bty = "n")

# Fitting variogram models to robust variogram estimator
vario.rob <- variog(coords = cbind(schoolsn$LONGITUDE,schoolsn$LATITUDE),data = res, estimator.type = "modulus", option = "bin",
vario.rob$n

dist.mat.earth <- rdist.earth(cbind(schoolsn$LONGITUDE,schoolsn$LATITUDE), miles = TRUE)
hmax= max(dist.mat.earth)/2
vario.rob.t <- vgram(loc=cbind(schoolsn$LONGITUDE,schoolsn$LATITUDE), y=res,breaks = seq(0,hmax,l=20),lon.lat=TRUE)

vario.rob$u <- vario.rob.t$centers
vario.rob$v <- vario.rob.t$stats["mean",]
vario.rob$sd <- vario.rob.t$stats["Std.Dev.",]
vario.rob$n <- vario.rob.t$stats["N",]


par(mgp = c(2,1,0))
plot(vario.rob$u,vario.rob$v,pch = 19,xlab = "lag (|h|)",ylab = expression(paste(gamma,"(|h|)")),main = "Emprical Variogram of M

expon.r <- variofit(vario.rob, ini.cov.pars = c(30,0.45),
                    cov.model = "exponential", fix.nugget = FALSE,
                    nugget = 130, weights = "cressie")
mat.r <- variofit(vario.rob, ini.cov.pars = c(30,0.45),
                  cov.model = "matern", fix.nugget = FALSE,
                  nugget = 130, weights = "cressie",
                  fix.kappa = FALSE, kappa = 0.5)
sph.r <- variofit(vario.rob, ini.cov.pars = c(25,7),
                  cov.model = "spherical", fix.nugget = FALSE,
                  nugget = 130, weights = "cressie")
cauch.r <- variofit(vario.rob, ini.cov.pars = c(30,0.45),
                    cov.model = "cauchy", fix.nugget = FALSE,
                    nugget = 130, weights = "cressie",
                    fix.kappa = FALSE, kappa = 0.5)

cauch.r$value
sqrt(sph.r$nugget)
```

```
sph.r$cov.pars
sph.r$kappa
sqrt(mat.r$cov.pars)


par(mar = c(4,4,3,3))
plot(vario.rob$u,vario.rob$v,pch = 19,xlab = "lag (|h|)",ylab = expression(paste(gamma,"(|h|)")),main = "Emprical Variogram of M
lines(cauch.r, col="black", lwd = 2, lty = 1)
lines(expon.r,col = "red",lwd = 2, lty = 3)
lines(mat.r, col="blue",lwd = 2, lty = 2)
lines(sph.r, col="orange", lwd = 2, lty = 4)
legend(0.02,170,legend = c("Exponential",
                           "Matern",
                           "Spherical",
                           "Cauchy"),
       col=c("red","blue","orange","black"),lty=c(3,2,4,1),lwd=c(2,2,2,2),bty = "n",y.intersp = 1.25)

# MLE Parameter fitting
library(data.table)
library(fields)
library(geoR)

res <- fread("residuals.csv")
res.res <- res$RESID

schools <- fread('CO_Schools_Data_Final.csv',drop=1)

schoolsn <- schools[!is.na(schools$PERFORMANCE)]
schoolsn <- schoolsn[!is.na(schoolsn$STR)]
schoolsn <- schoolsn[!is.na(schoolsn$FRL)]

dist.mat <- rdist.earth(data.frame(lon = schoolsn$LONGITUDE,lat = schoolsn$LATITUDE),miles=TRUE)

n <- nrow(schoolsn)

X1 <- cbind(1,schoolsn$FRL,schoolsn$STR)
ell <- function(p){
  # p = log(tau,sigma,a,nu)
  Sigma <- diag(p[1]^2,n) + p[2]^2*Matern(dist.mat,range=p[3],nu=0.5)
  Sigma.i <- solve(Sigma)

  # negative log-likelihood value
  out <- 0.5 * n * log(2*pi) + 0.5 * determinant(Sigma,log=TRUE)$modulus +
    0.5 * t(schoolsn$PERFORMANCE) %*%
    (Sigma.i - Sigma.i %*% X1 %*% solve(t(X1) %*% Sigma.i %*% X1) %*% t(X1) %*% Sigma.i) %*%
    schoolsn$PERFORMANCE

  print(out)
}
mod1 <- optim(par=c(11,5,7),fn=ell,hessian=TRUE)
pars <- mod1$par
pars
Sigmaf <- diag(pars[1]^2,n) + pars[2]^2*Matern(dist.mat,range=pars[3],nu=0.5)
Sigmaf.i <- solve(Sigmaf)
beta <- solve(t(X1) %*% Sigmaf.i %*% X1) %*% t(X1) %*% Sigmaf.i %*% schoolsn$PERFORMANCE
beta

#aic
2*mod1$value + 2*(6+1)
#bic
2*mod1$value + (6+1)*log(n)

#cov parameter errors
I.inv <- solve(mod1$hessian)
SEs <- sqrt(diag(I.inv))
SEs

#beta errors
```

```
BSEs <- sqrt(diag(solve(t(X1) %*% Sigmaf.i %*% X1)))
BSEs

### Pure Matern
ell2 <- function(p){
  # p = log(tau,sigma,a,nu)
  Sigma <- diag(p[1]^2,n) + p[2]^2*Matern(dist.mat,range=p[3],nu=p[4])
  Sigma.i <- solve(Sigma)

  # negative log-likelihood value
  out <- 0.5 * n * log(2*pi) + 0.5 * determinant(Sigma,log=TRUE)$modulus +
    0.5 * t(schoolsn$PERFORMANCE) %*%
    (Sigma.i - Sigma.i %*% X1 %*% solve(t(X1) %*% Sigma.i %*% X1) %*% t(X1) %*% Sigma.i) %*%
    schoolsn$PERFORMANCE

  print(out)
}
mod2 <- optim(par=c(11.68,5.88,7,0.47),fn=ell2,hessian=TRUE)
pars2 <- mod2$par
pars2
Sigmaf2 <- diag(pars2[1]^2,n) + pars2[2]^2*Matern(dist.mat,range=pars2[3],nu=pars2[4])
Sigmaf2.i <- solve(Sigmaf2)
beta2 <- solve(t(X1) %*% Sigmaf2.i %*% X1) %*% t(X1) %*% Sigmaf2.i %*% schoolsn$PERFORMANCE
beta2
#aic
2*mod2$value + 2*(7+1)
#bic
2*mod2$value + (7+1)*log(n)

I.inv2 <- solve(mod2$hessian)
SEs2 <- sqrt(diag(I.inv2))
SEs2

#beta errors
BSEs2 <- sqrt(diag(solve(t(X1) %*% Sigmaf2.i %*% X1)))
BSEs2

#Kriging our model residuals
rm(list=ls())
library(fields)
library(data.table)

schools <- fread('CO_Schools_Data_Final.csv',drop=1)

schoolsn <- schools[!is.na(schools$PERFORMANCE)]
schoolsn <- schoolsn[!is.na(schoolsn$STR)]
schoolsn <- schoolsn[!is.na(schoolsn$FRL)]

tau2 <- 11.202992^2
sigma2 <- 5.787580^2
lambda <- 9.839605
B0 <- 81.0335820
B1 <- -0.3432731
B2 <- -0.1730419

res <- schoolsn$PERFORMANCE - (B0 + B1*schoolsn$FRL + B2*schoolsn$STR)
res <- as.matrix(res)
xy <- cbind(schoolsn$LONGITUDE, schoolsn$LATITUDE)


dist.mat <- rdist.earth(xy, miles = TRUE)
dim(dist.mat)
n <- dim(dist.mat)[1]

Sigma <- sigma2*Matern(dist.mat,range=lambda,nu=0.5)
Sigma.i <- solve(Sigma)

# Simple Kriging on residuals
```

```
# Simga_0 matrix calculation
x <- seq(min(xy[,1]),max(xy[,1]),l=100)
y <- seq(min(xy[,2]),max(xy[,2]),l=100)
xy_topred <- expand.grid(x,y)
dist.mat_0 <- rdist.earth(xy_topred,xy, miles = TRUE)
dim(dist.mat_0)
Sigma_0 <- sigma2*Matern(dist.mat_0,range=lambda,nu=0.5)
dim(Sigma_0)

fhat <- Sigma_0 %*% Sigma.i %*% res

#uncertainity
sk2 <- vector("numeric", 10000)
for(i in 1:10000){
  sk2[i] <- sigma2+tau^2 - (t(as.matrix(Sigma_0[i,])) %*% Sigma.i %*% as.matrix(Sigma_0[i,]))
}
sk <- sqrt(sk2)

fhat.mat <- matrix(fhat,100,100)
sk.mat <- matrix(sk,100,100)

par(mfrow = c(1,1), mgp = c(2.5,1,0), mar = c(4,4,3,3))
image.plot(x,y,fhat.mat, xlab = "Longitude", ylab = "Latitude", main = "Kriged Residuals")
mtext("Residual", 4, line=0.5)

image.plot(x,y,sk.mat, xlab = "Longitude", ylab = "Latitude", main = "Kriged Residuals Standard Error")
mtext("Standard Error", 4, line=0.5)

par(mfrow = c(1,2),mar = c(4,2,4,5))
quilt.plot(xy_topred[1],xy_topred[2],fhat)
#quilt.plot(xy_topred[1],xy_topred[2],fhat,nx=100,ny=100)
quilt.plot(xy[,1],xy[,2],res)

# universal kriging
X <- as.matrix(cbind(1,schoolsn$FRL,schoolsn$STR))
X.t <- t(X)

#places to predict
xytopred <- matrix(c(-104.4,39.75,-108.5,41,-105,38,-107,39.5),nrow = 4, ncol = 2,byrow = TRUE)

dist.mat.univ <- rdist.earth(xytopred, xy, miles=TRUE)

x_0 <- matrix(c(1,20,20, 1,70,5, 1,10,10, 1,50,50),nrow = 4, ncol = 3, byrow = TRUE)
x_0.t <- t(x_0)

Sigma_0 <- sigma2*Matern(dist.mat.univ,range=lambda,nu=0.5)
Sigma_0.t <- t(Sigma_0)

fhat.univ <- (Sigma_0 + (x_0 - (Sigma_0 %*% Sigma.i %*% X)) %*%
                solve(X.t %*% Sigma.i %*% X) %*% X.t) %*% Sigma.i %*% schoolsn$PERFORMANCE
fhat.univ

uncertainty.univ <- sigma2 + tau2 - (Sigma_0 %*% Sigma.i %*% Sigma_0.t) +
                              (t(t(x_0) - (X.t %*% Sigma.i %*% Sigma_0.t)) %*%
                                solve(X.t %*% Sigma.i %*% X) %*%
                                 (t(x_0) - (X.t %*% Sigma.i %*% Sigma_0.t)))
uk <- sqrt(diag(uncertainty.univ))
uk
```